DIGITAL SIGNAL PROCESSING

# LECTURE (10)

# Digital Signal Processing Applications

Assist. Prof. Amr E. Mohamed

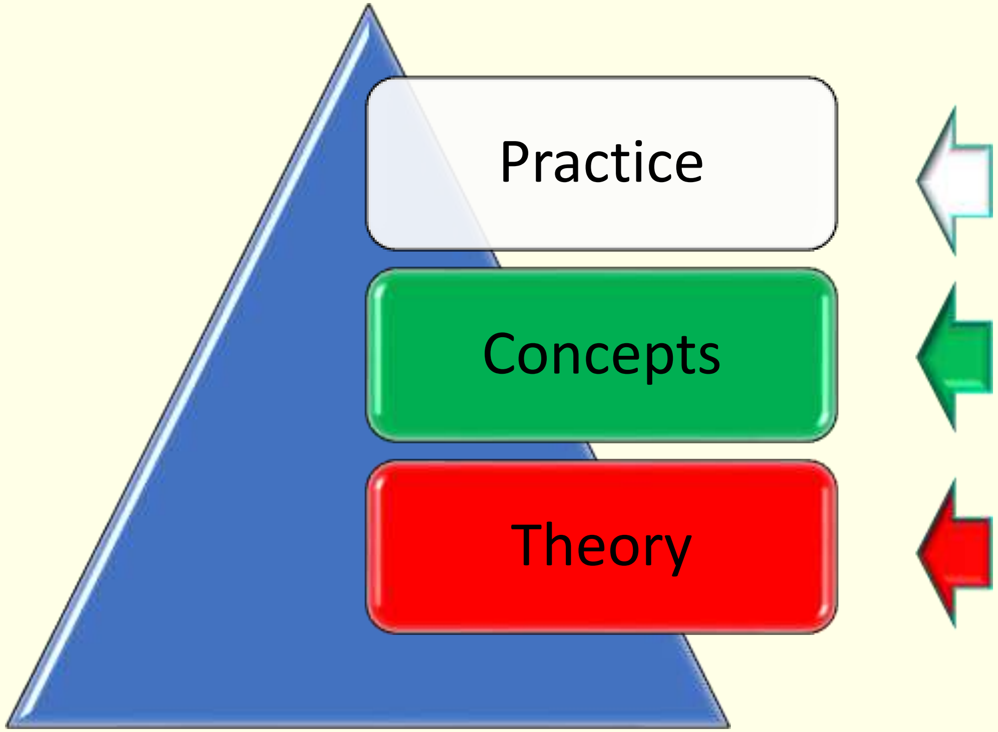
لما نصدق اننا نِقْدَر نِقْدَر ،،،

# Agenda

- Introduction
- Digital Spectrum Analyzer
- Speech Processing
- Radar

# Introduction

- Three levels of understanding are required to appreciate the technological advancement of DSP
  - Theoretical
  - Conceptual
  - Practical




Ability to implement theory and concepts in working code (MATLAB, C, C++, JAVA)

Basic understanding of how theory is applied

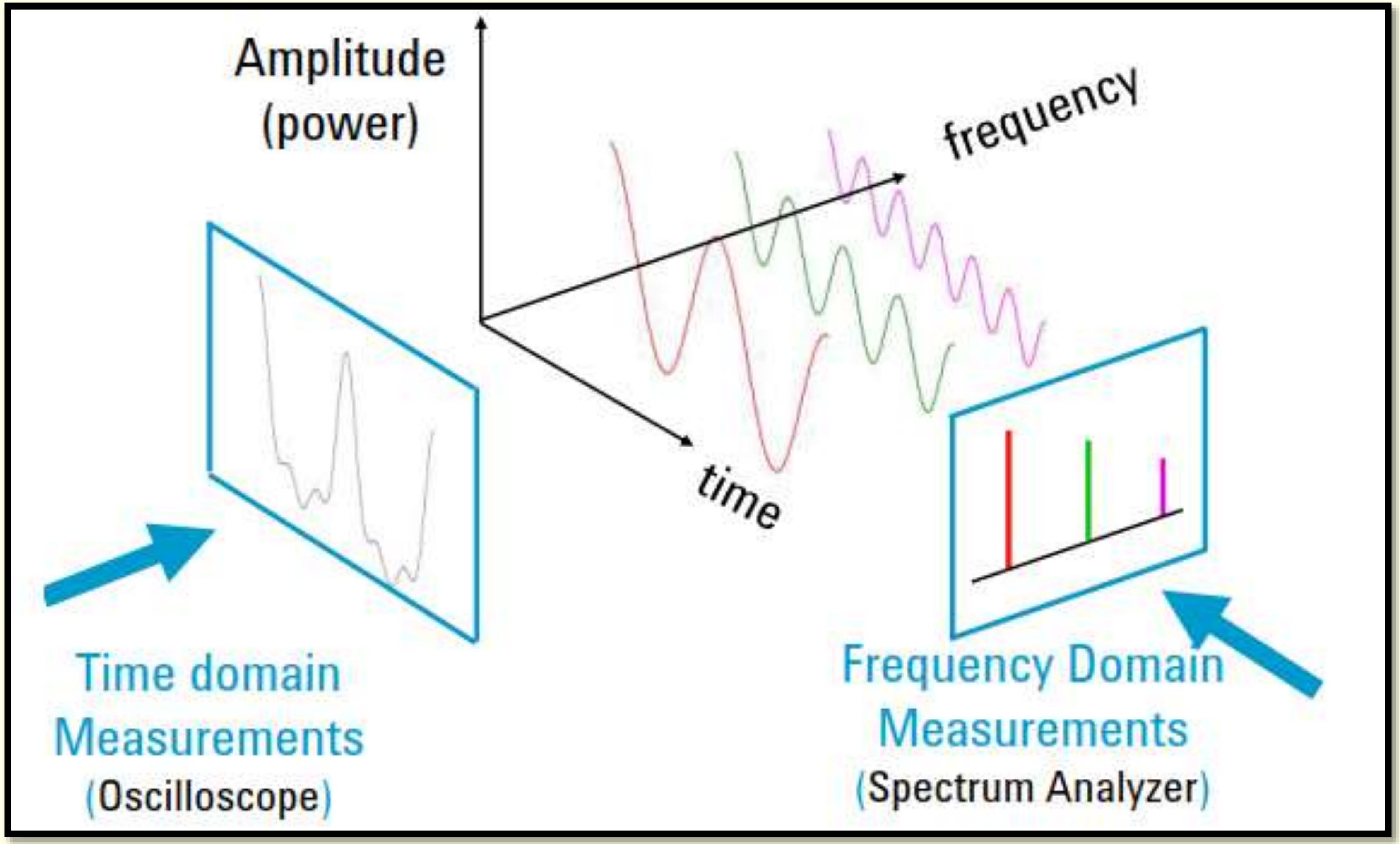
Mathematics, derivations, signal processing

# Digital Spectrum Analyzer

# Topics

- Time Domain & Frequency Domain.
- Fourier Analysis.
- Decibel.
- Spectrum Analyzer.
  - Filter Bank Spectrum Analyzer.
  - Swept Spectrum Analyzer.
  - FFT Spectrum Analyzer.
  - Real-Time Spectrum Analyzer.

# Time and Frequency Measurements

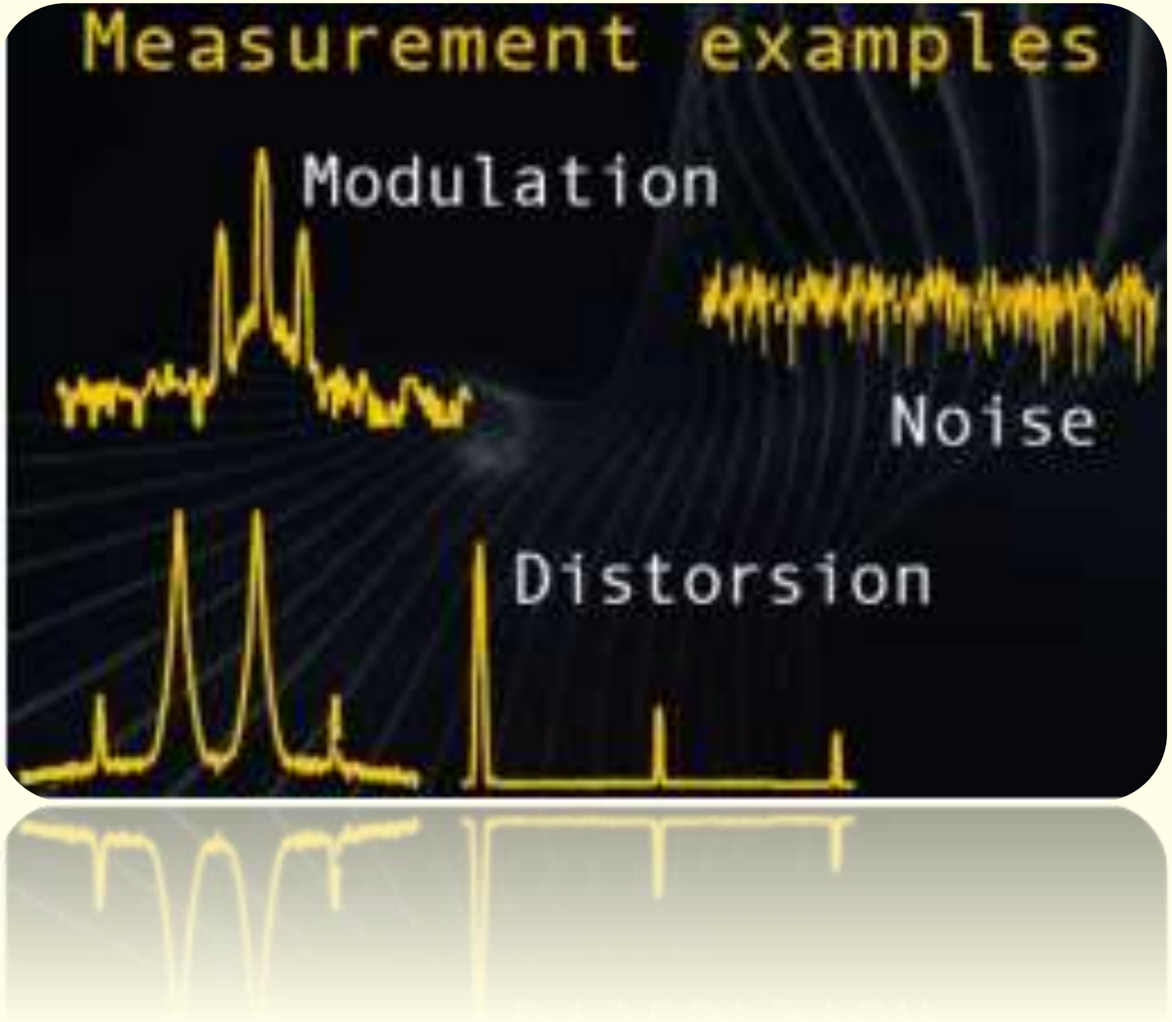


# Time and Frequency Measurements

- A time domain graph shows how the signal amplitude changes over time.
- To fully understand the performance of your device/system, you will also want to analyze the signal(s) in the frequency-domain.
- The spectrum analyzer is to the frequency domain as the oscilloscope is to the time domain.

# Time and Frequency Measurements

- Some fields of application
  - Telecommunications.
  - Analog and Power Electronics.
  - Electromagnetic.
  - Electrical Systems and Machines
  - Bioengineering



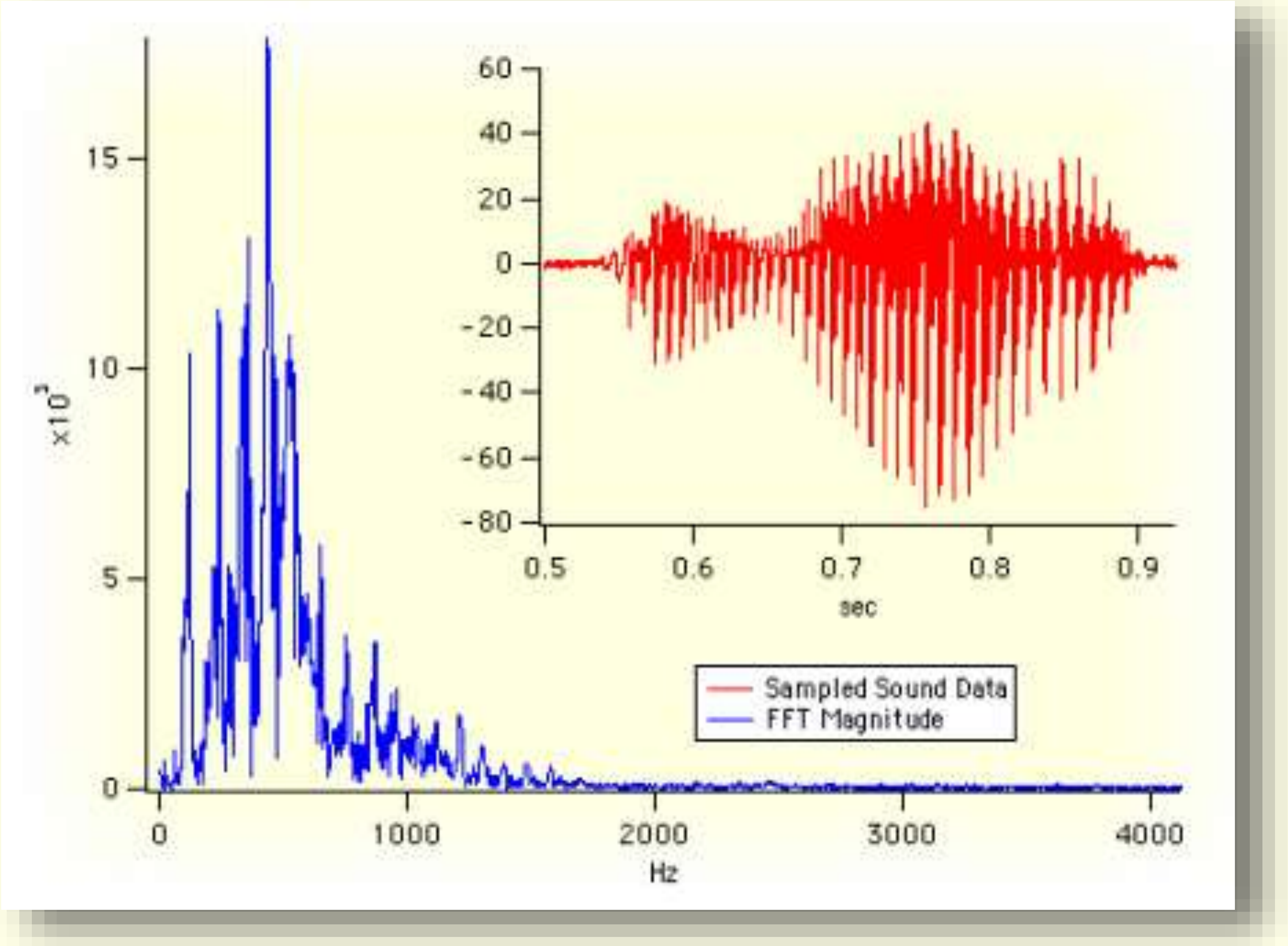
# Frequency Response

**Fourier analysis**

is used to get the frequency response of time domain signal.

- Fourier analysis:
  - Fourier Series.
  - Fourier Transform.
  - Discrete-Time Fourier Transform (DTFT).
  - Discrete Fourier Transform (DFT).
  - **FAST FOURIER TRANSFORM (FFT)**

# Fourier Transform



# IMPORTANCE AND USAGE OF DECIBEL

## Decibel definition

Ratio between two power values:

$$A_{dB} = 10 \cdot log(\frac{P_2}{P_1})$$

$P_2$ is expressed in relative terms with respect to $P_1$.

Swapping $P_2$ and $P_1$ changes the sign of A( dB).

# IMPORTANCE AND USAGE OF DECIBEL

## Role of resistances

If the powers refer to a couple of voltages applied to two resistances:

$$A_{dB} = 10 \cdot log(\frac{V_2^2/R_2}{V_1^2/R_1}) = 20 \cdot log(\frac{V_2}{V_1}) + 10 \cdot log(\frac{R_1}{R_2})$$

$V_2$ and $V_1$ are the rms voltages applied to $R_2$ and $R_1$ respectively.

If $R_2 = R_1$

$$A_{dB} = 20 \cdot log(\frac{V_2}{V_1})$$

# IMPORTANCE AND USAGE OF DECIBEL

## Fundamentals rules

**Inversion**

$$A_{dB} = 10 \cdot log(Q) \quad \Rightarrow \quad -A_{dB} = 10 \cdot log(\frac{1}{Q})$$

**Addition**

$$10 \cdot log(Q_1) + 10 \cdot log(Q_2) = 10 \cdot log(Q_1 \cdot Q_2)$$

**Subtraction**

$$10 \cdot log(Q_1) - 10 \cdot log(Q_2) = 10 \cdot log(\frac{Q_1}{Q_2})$$

# IMPORTANCE AND USAGE OF DECIBEL

## Absolute values in Decibel: *dBm*

Choosing **1** $mW$ as reference power value

$$P_{dBm} = 10 \cdot log(P_{mW})$$

$P_{mW}$ is the power expressed in milliwatt, it is dependent of $R$

$$P_{dBm} = 20 \cdot log(\frac{V}{V_{REF}})$$

$V_{REF}$ is the rms voltage producing 1 $mW$ power:

$$V_{REF} = \sqrt{P \cdot R} = \sqrt{0.001 \cdot R}$$

# IMPORTANCE AND USAGE OF DECIBEL

**Absolute values in Decibel: *dBm***

for $R = 50\Omega \quad \Rightarrow \quad V_{REF} = 0.2236\ Volt$

$$P_{dBm} = 20 \cdot log(\frac{V_{rms}}{0.2236})$$

for $R = 75\Omega \quad \Rightarrow \quad V_{REF} = ??\ Volt$

$$P_{dBm} = ??$$

# IMPORTANCE AND USAGE OF DECIBEL

## Absolute values in Decibel: *dBV*

Choosing **1** $V$ rms as reference value

$$V_{dBV} = 20 \cdot log(V_{rms})$$

$V_{dBV}$ is <u>independent of $R$</u>

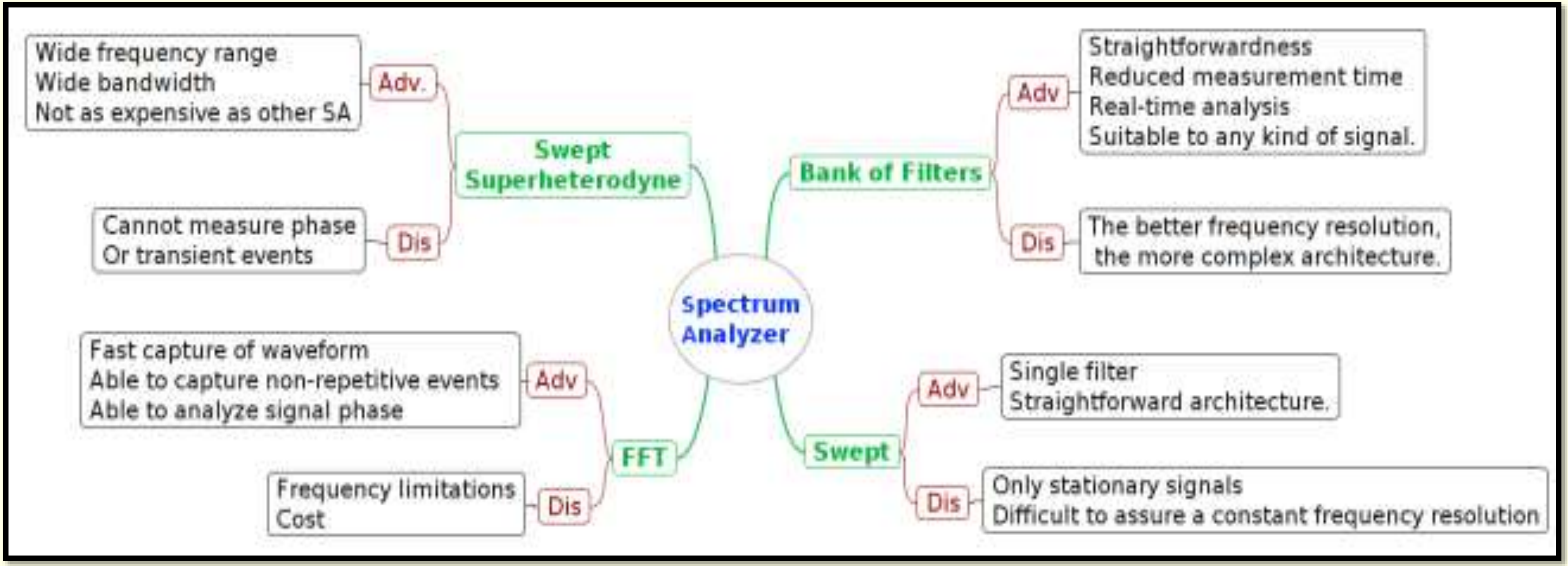
## $dBm/dBV$ conversion

$$P_{dBm} = V_{dBV} + 10 \cdot log(\frac{1}{0.001 \cdot R})$$

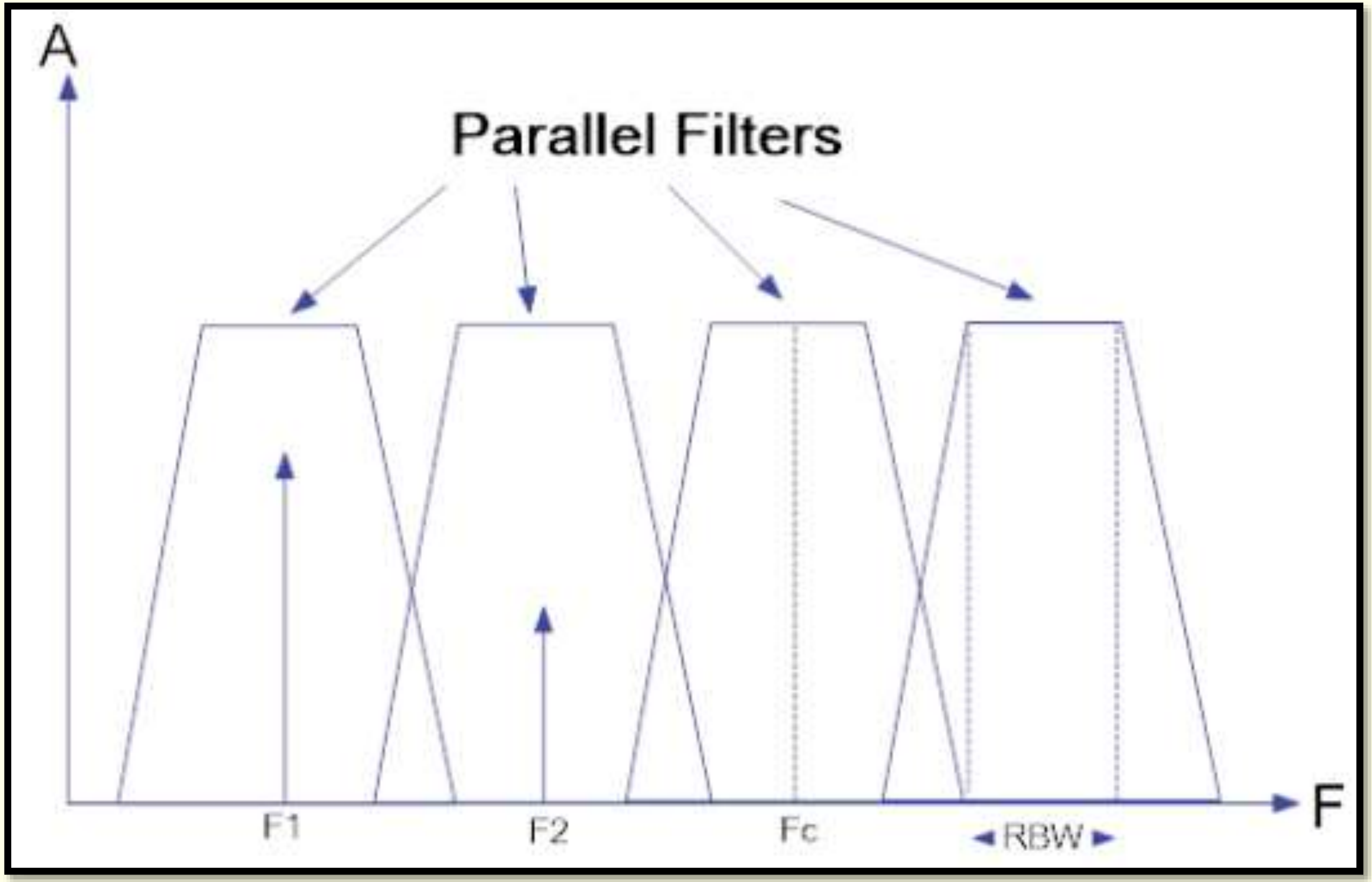
for $R = 50\Omega \quad \Rightarrow \quad P_{dBm} = V_{dBV} + 13.01$

for $R = 75\Omega \quad \Rightarrow \quad P_{dBm} = V_{dBV} + \ldots$

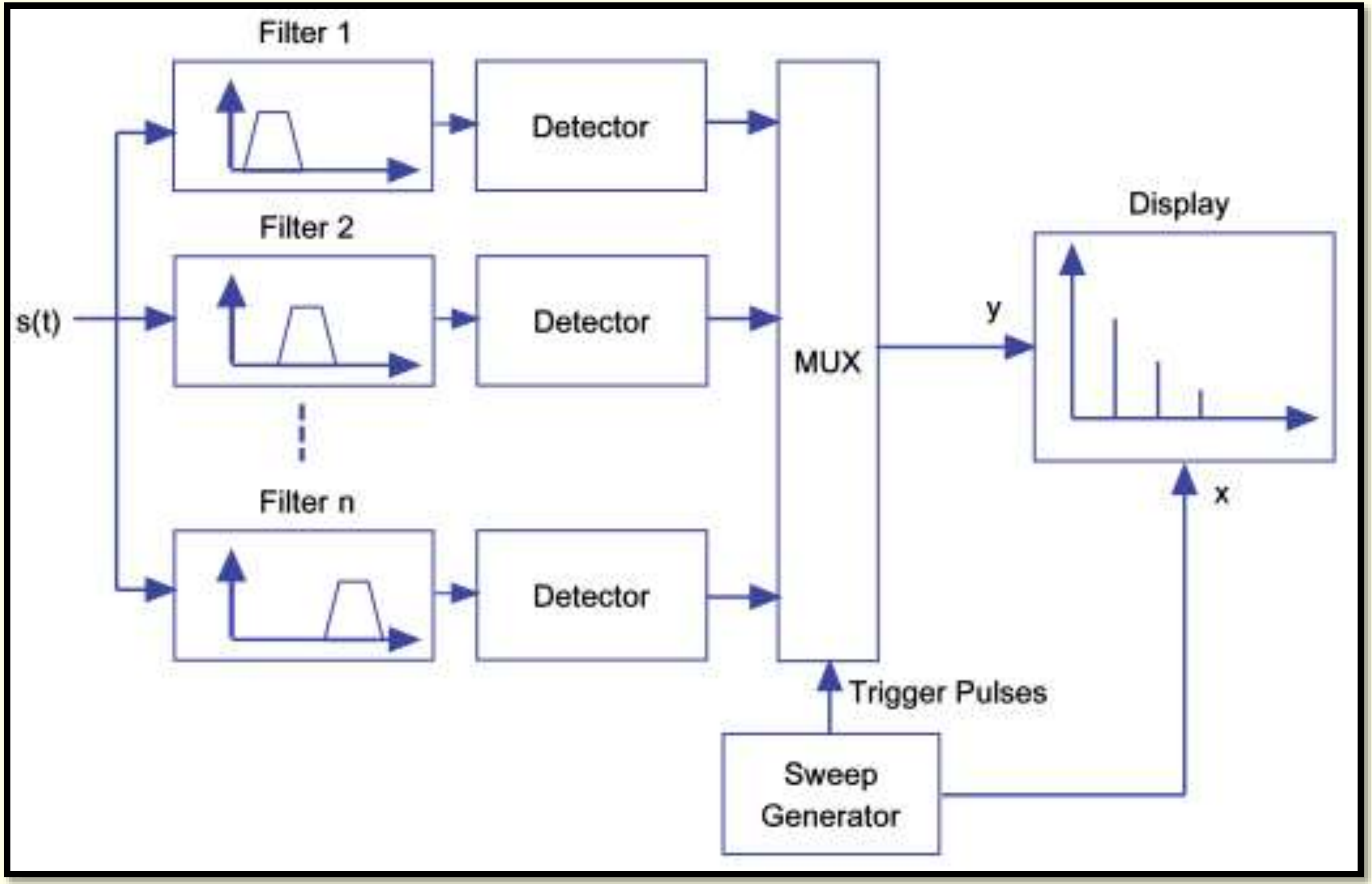
# Digital Spectrum Analyzer



# 1. Bank-of-filters Spectrum Analyzer



# 1. Bank-of-filters Spectrum Analyzer



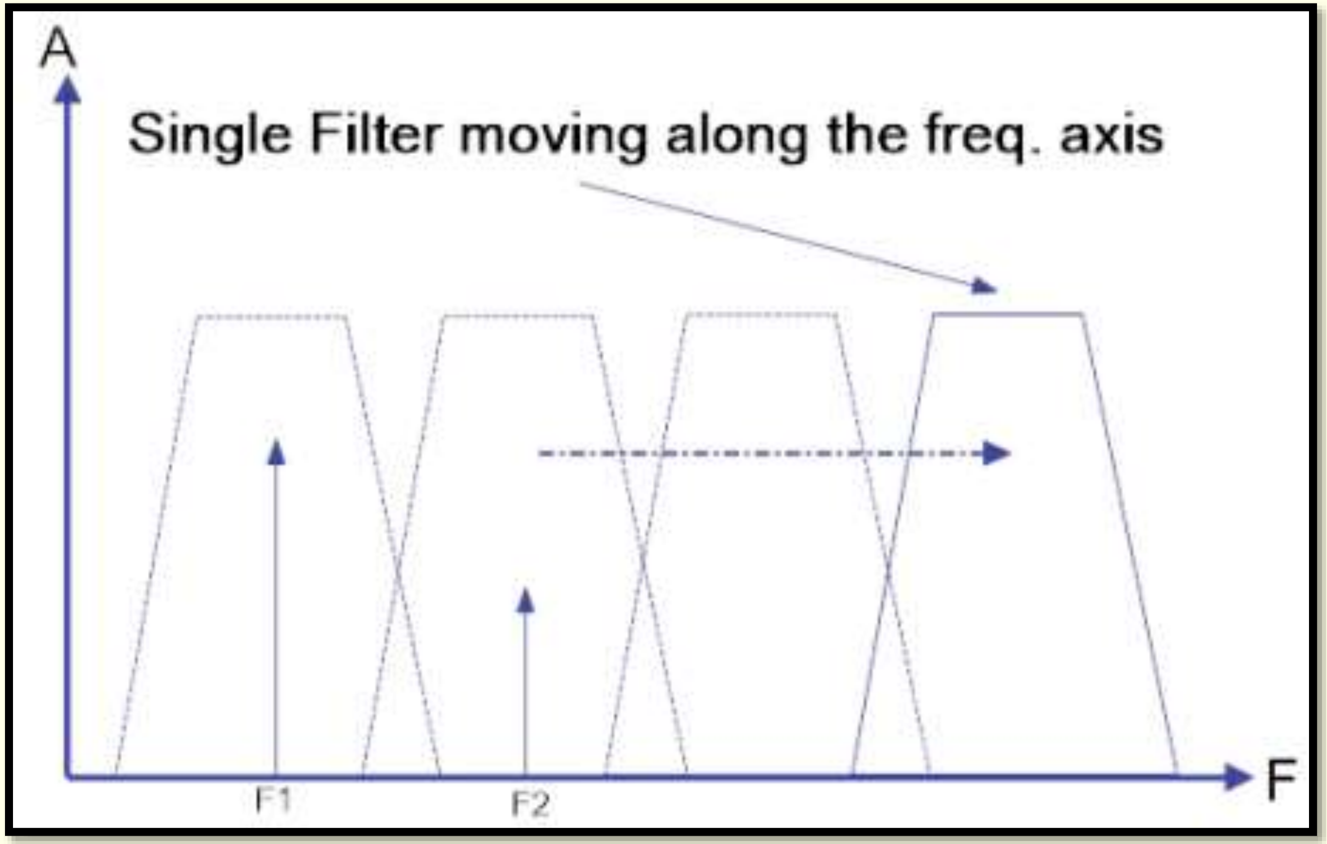
# 1. Bank-of-filters Spectrum Analyzer

- Main Design
  - Fixed pass-band filters
  - Minimum overlap between the frequency responses of two adjacent filters.
  - Detectors for measuring the ”sinusoidal” output of the filters.
- Frequency resolution
  - It is strictly dependent on RBW (resolution bandwidth) value
- Drawbacks Example:

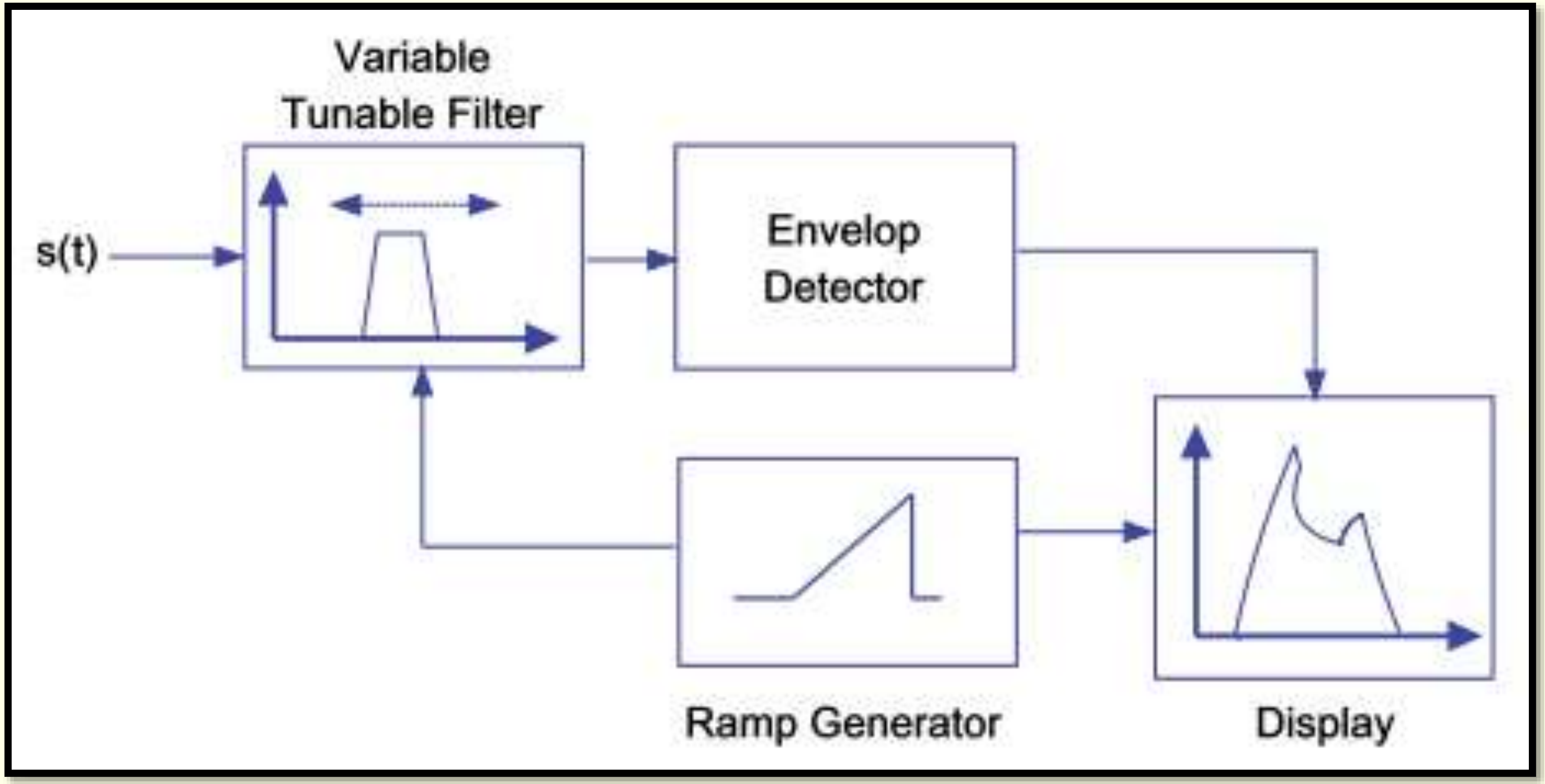
- Frequency interval = 0 Hz - 100 kHz
- Resolution = $\Delta f$ = 100 Hz
- Number of Filters = M $= \frac{100KHz}{100Hz} = 1000$ ☹☹☹

# 2. Swept Spectrum Analyzers

- ❑ Sequential Spectrum Analyzer.
- ❑ Variable tuning filter spectrum analyzer.

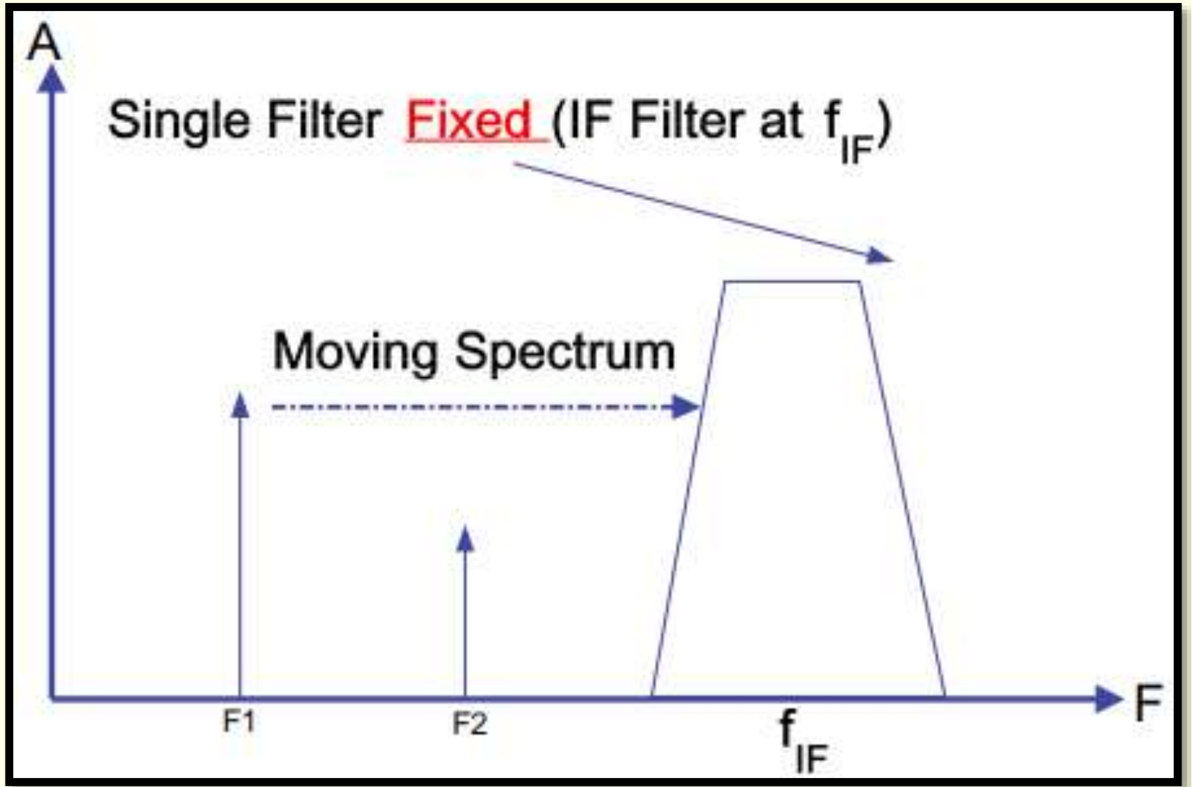


# 2. Swept Spectrum Analyzers

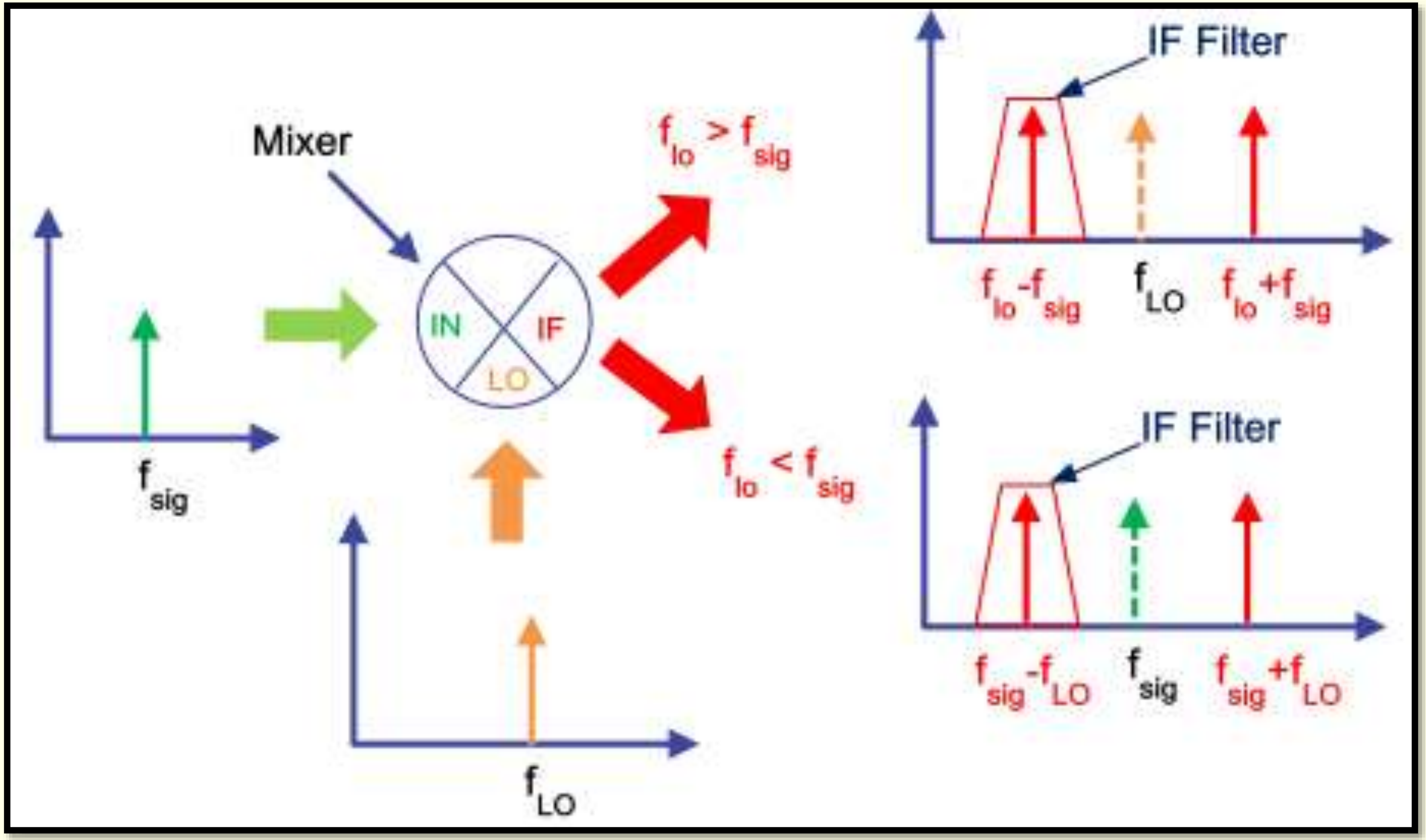


# 3. Swept Superheterodyne Spectrum Analyzers

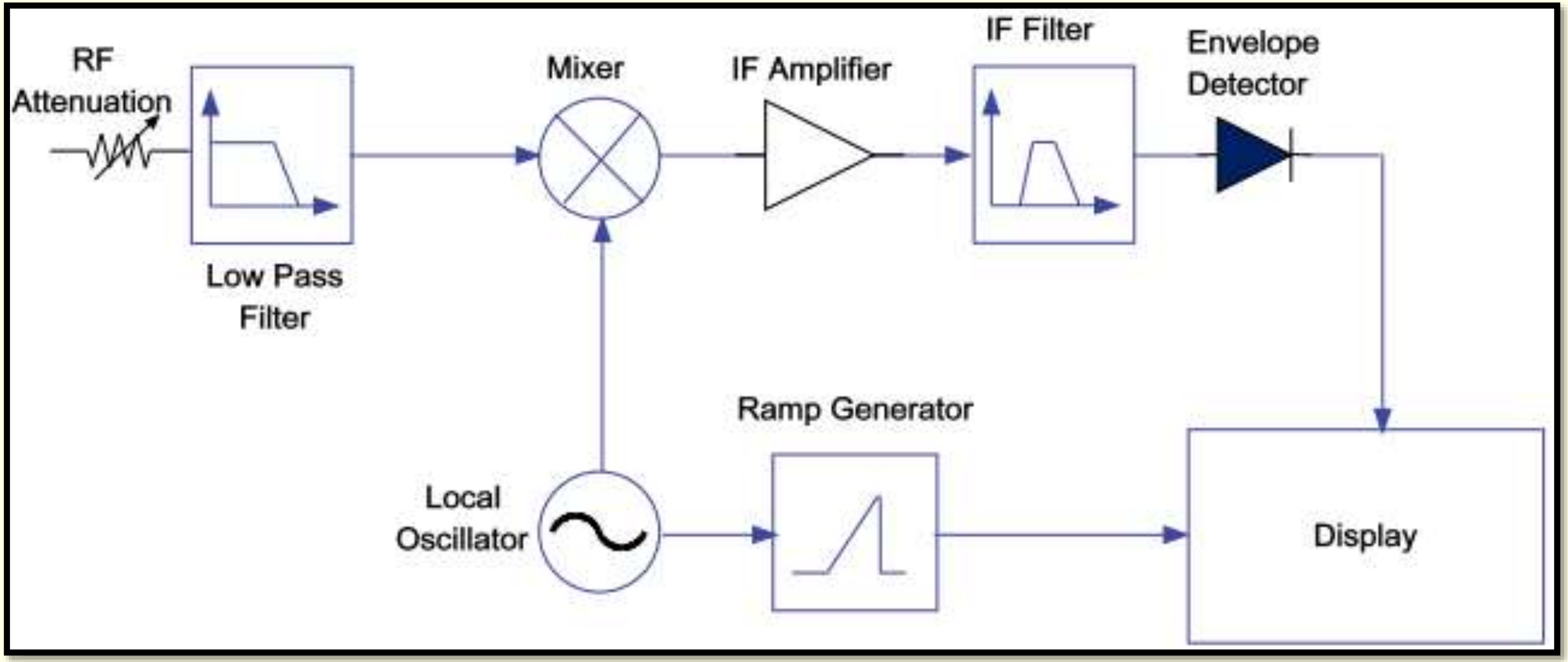
- Main design choices
  - Single pass-band filter with constant central frequency
  - Mixer implementing the superheterodyne technique
  - Detector for measuring the envelope of the filter output signal



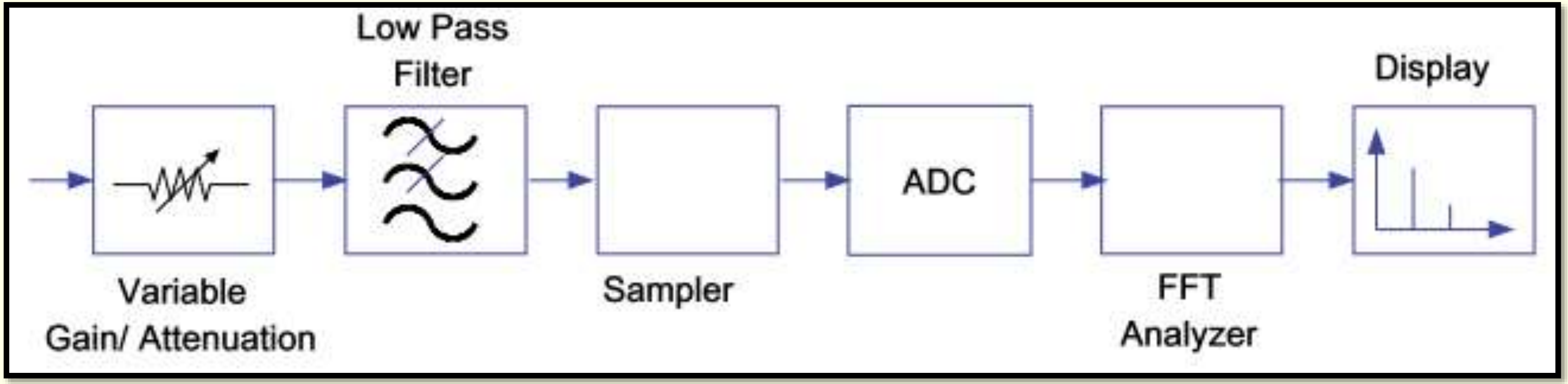
# 3. Swept Superheterodyne Spectrum Analyzers



# 3. Swept Superheterodyne Spectrum Analyzers



# 4. FFT spectrum analyzer



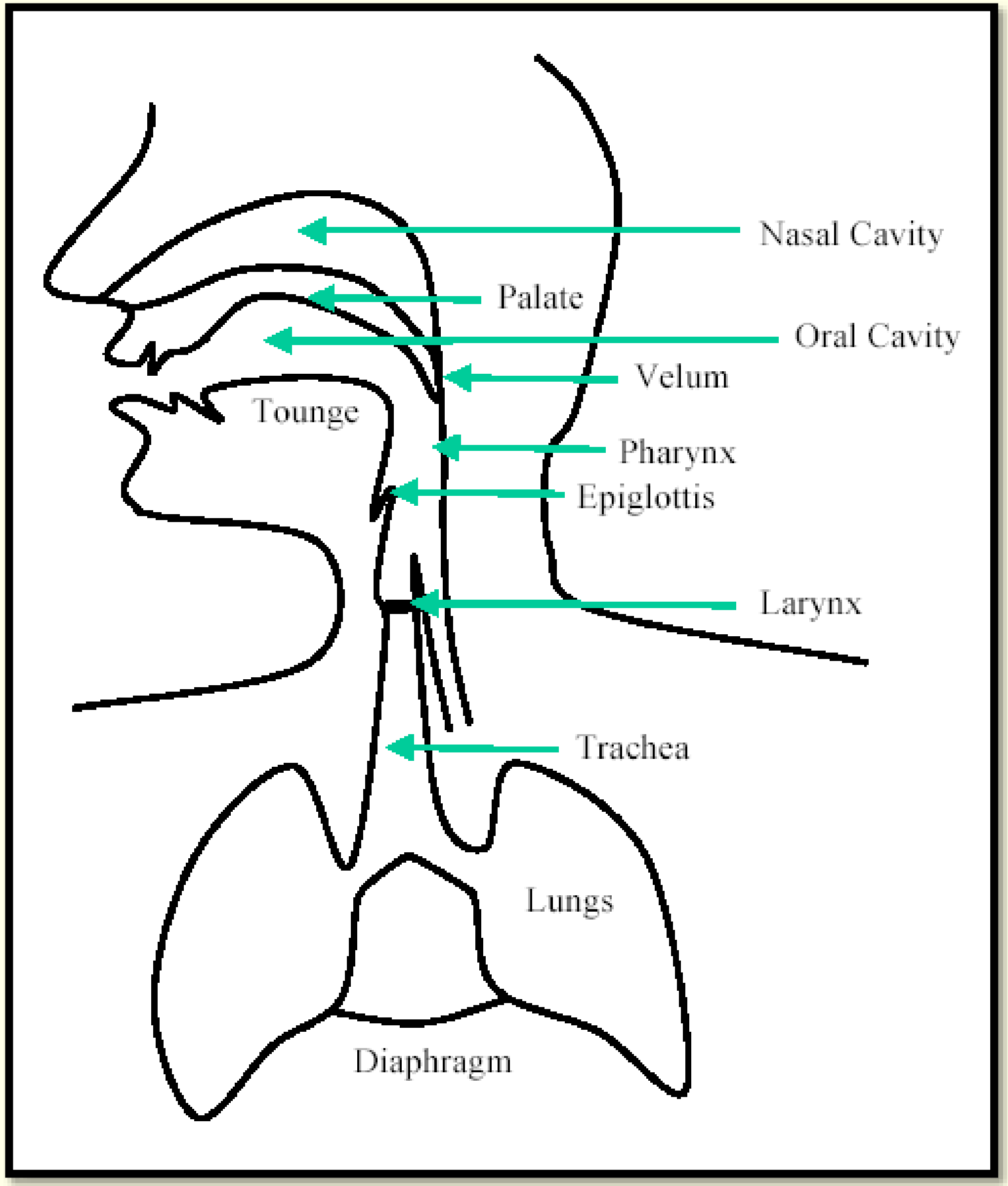
# Speech Processing

# Speech Generation and Perception

- The study of the anatomy of the organs of speech is required as a background for articulatory and acoustic phonetics.

- An understanding of hearing and perception is needed in the field of both speech synthesis and speech enhancement and is useful in the field of automatic speech recognition.

# Schematic diagram of the human speech production



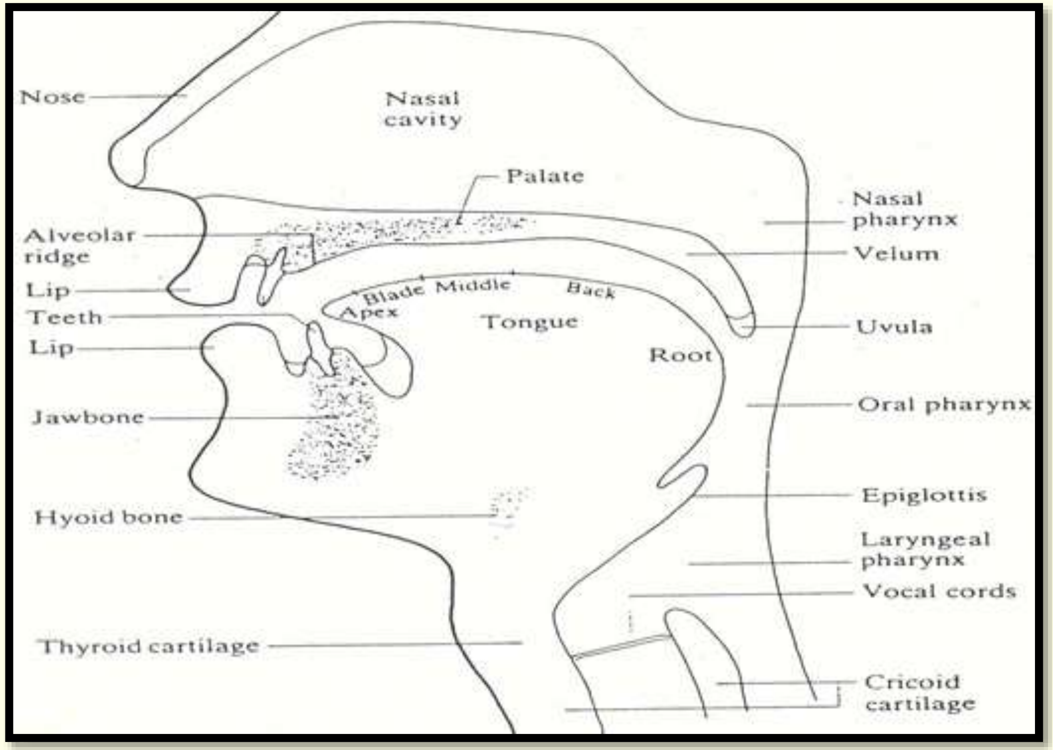
# Organs of Speech

- Lungs and trachea:
  - source of air during speech.
  - The vocal organs work by using compressed air; this is supplied by the lungs and delivered to the system by way of the trachea.
  - These organs also control the loudness of the resulting speech.

- The Larynx:
  - This is a complicated system of cartilages and muscle containing and controlling the vocal cords.
  - The vocal cords are composed of twin infoldings of mucous membrane stretched horizontally, from back to front, across the larynx.
  - They vibrate, modulating the flow of air being expelled from the lungs during phonation.
  - Open when breathing and vibrating for speech or singing, the folds are controlled via the vagus nerve.

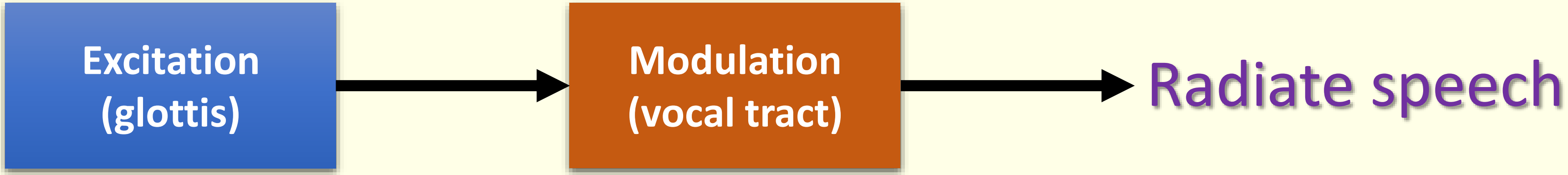
# Organs of Speech

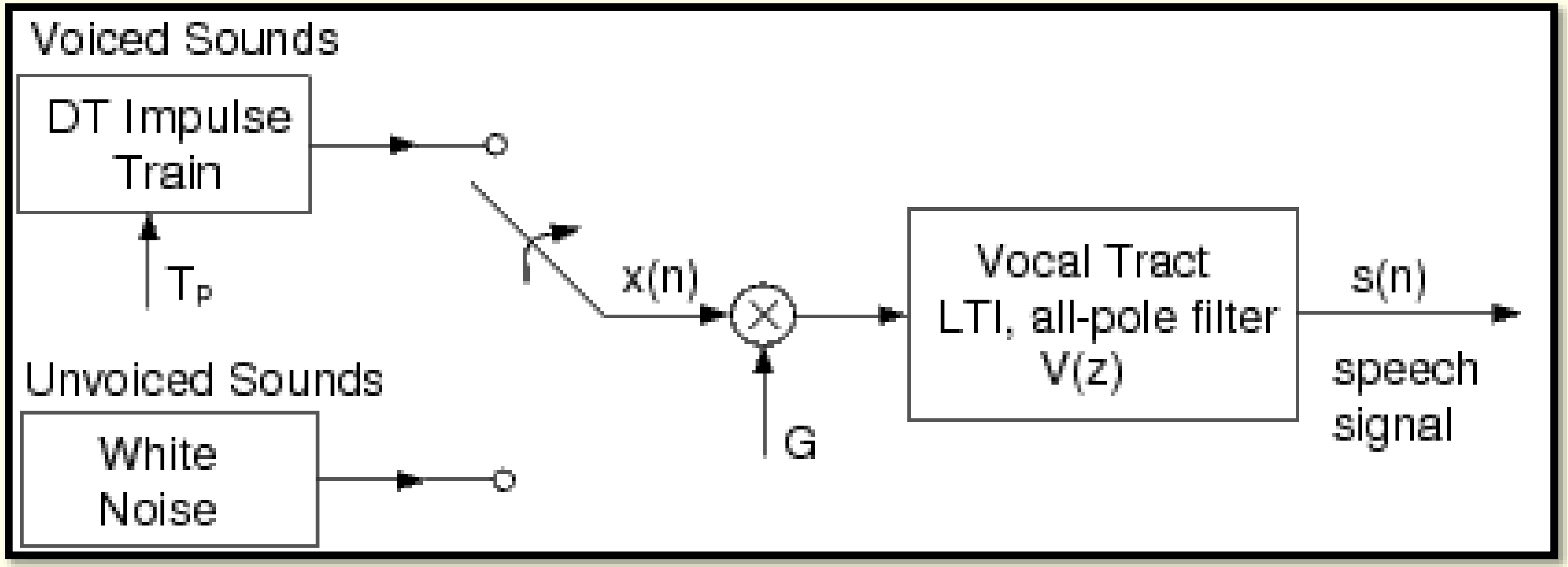
- **The Vocal Tract:**
  - The **vocal tract** is the cavity in human beings and in animals where sound that is produced at the sound source is filtered.
    - **Oral cavity**: Forward of the velum and bounded by lips, tongue and palate
    - **Nasal cavity**: Above the palate and extending from the pharynx to the nostrils



# A General Discrete-Time Model For Speech Production

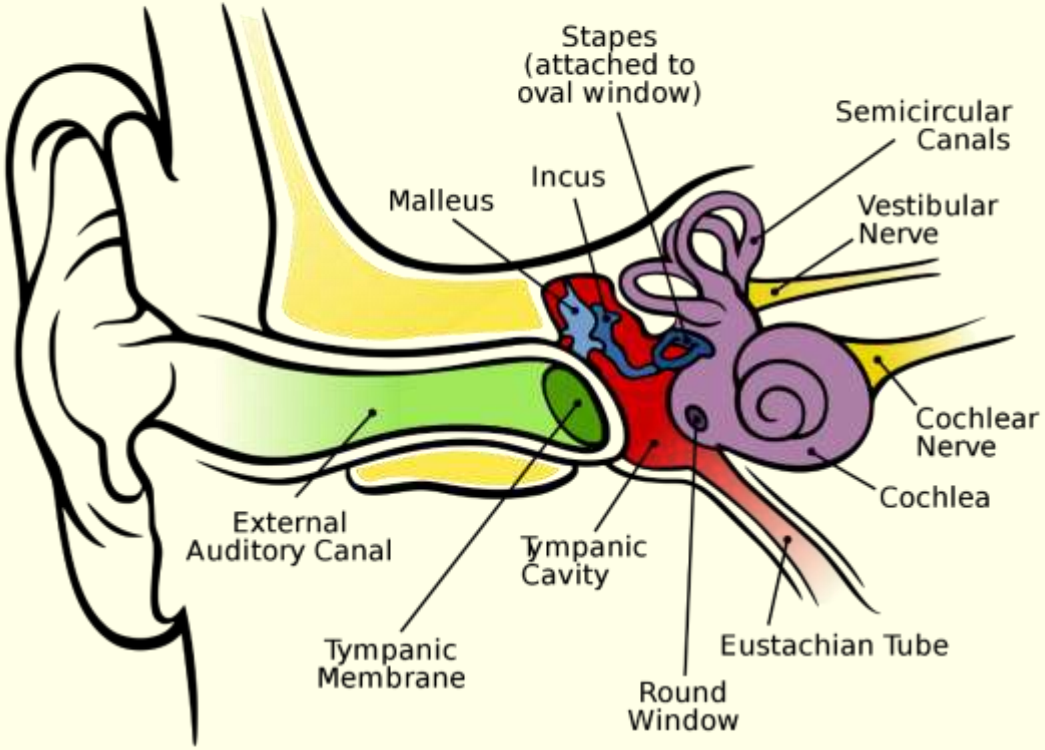
- The operation of the system is divided into two functions :
  - Excitation
  - Modulation



# Hearing and perception

- Hearing is a process which sound is received and convert into nerve impulse
- Perception is the post-processing within the brain by which the sounds heard are interpreted and given meaning



# What is the difference between consonant and vowel?

- A consonant is a sound that is made with the air stopping once or more during the vocalization. That means that at some point, the sound is stopped by your teeth, tongue, lips, or constriction of the vocal cords.
  - Voiced
  - Unvoiced
- A vowel is a sound that is made with the mouth and throat not closing at any point.
  - Tongue high or low
  - Tongue front or back
  - Lips rounded or unrounded

# Voiced and Unvoiced sounds

- There are two mutually exclusive ways excitation functions to model voiced and unvoiced speech sounds.
- For a short time-basis analysis:
  - voiced speech is considered periodic with a fundamental frequency of $f_0$, and a pitch period of $1/f_0$, which depends on the speaker. Hence, Voiced speech is generated by exciting the all pole filter model by a periodic impulse train.
- On the other hand, unvoiced sounds are generated by exciting the all-pole filter by the output of a random noise generator

# Voiced/Unvoiced

- The fundamental difference between these two types of speech sounds comes from the following:
  - the way they are produced.
  - The vibrations of the vocal cords produce voiced sounds.
  - The rate at which the vocal cords vibrate dictates the pitch of the sound.
  - On the other hand, unvoiced sounds do not rely on the vibration of the vocal cords.
  - The unvoiced sounds are created by the constriction of the vocal tract.
  - The vocal cords remain open and the constrictions of the vocal tract force air out to produce the unvoiced sounds

- Given a short segment of a speech signal, lets say about $20ms$ or $160$ samples at a sampling rate $8KHz$, the speech encoder at the transmitter must determine the proper excitation function, the pitch period for voiced speech, the gain, and the coefficients

# Phonemics (phonemes):

- A phoneme is the smallest sound unit in a given language that is sufficient to differentiate one word from another
- English Phonemes

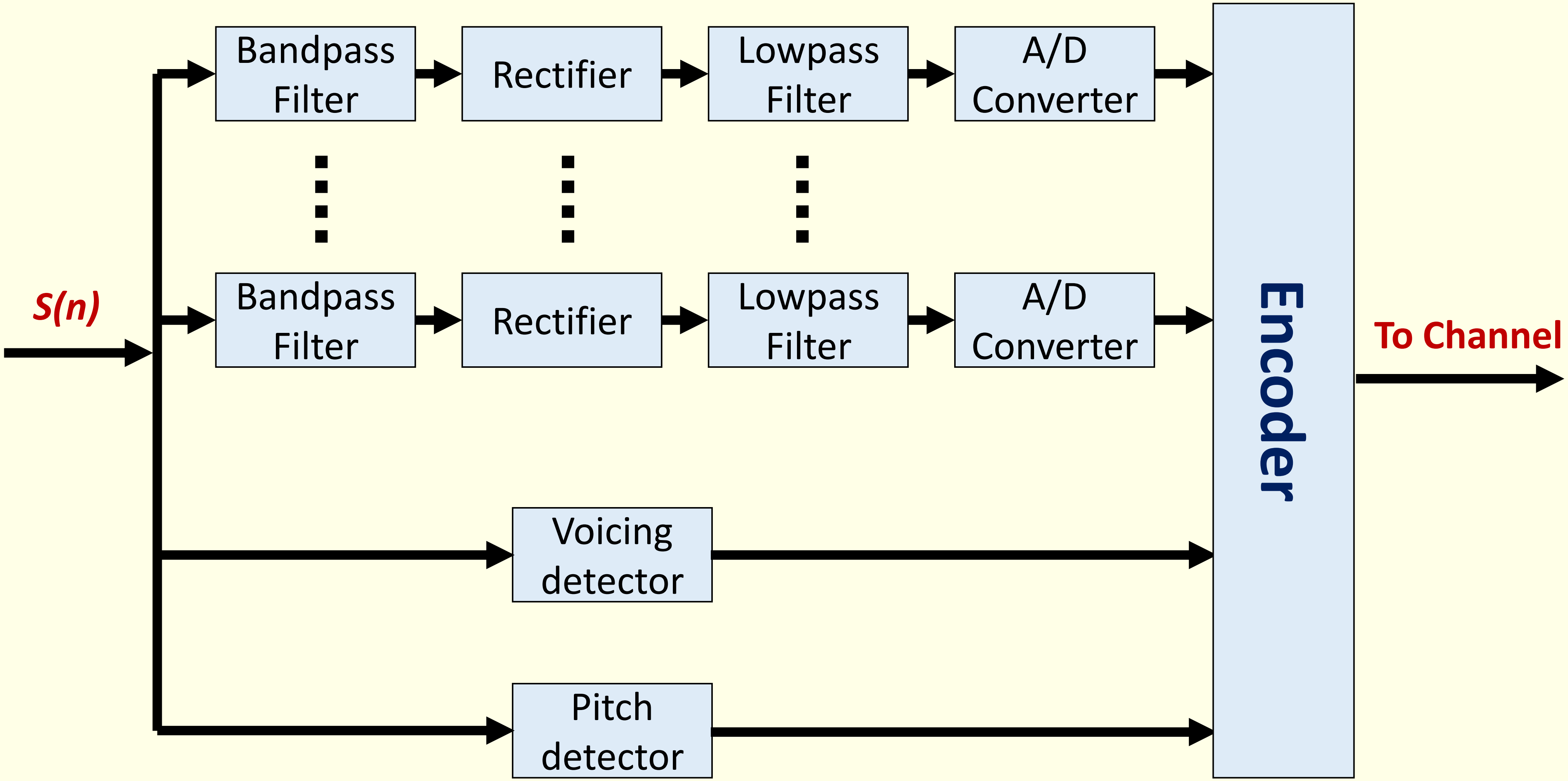
| | |
|---|---|
| **Vowels** | uw ux uh ah ax ah-h aa ao ae eh<br>ih ix ey iy ay ow aw oy er axr el |
| **Semi-vowels** | y r l el w |
| **Fricatives** | jh ch s sh z zh f th v dh |
| **Nasals** | m n ng em en eng nx |
| **Stops** | b d g p t k dx q<br>bcl dcl gcl pcl tcl kcl |
| **Aspiration** | hv hh |

# Audio and Speech Compression

# The Channel Vocoder - analyzer

- The channel vocoder employs a bank of bandpass filters,
  - Each having a bandwidth between 100HZ and 300HZ.
  - Typically, 16-20 linear phase FIR filter are used.
- The output of each filter is rectified and lowpass filtered.
  - The bandwidth of the lowpass filter is selected to match the time variations in the characteristics of the vocal tract.
- For measurement of the spectral magnitudes, a voicing detector and a pitch estimator are included in the speech analysis.

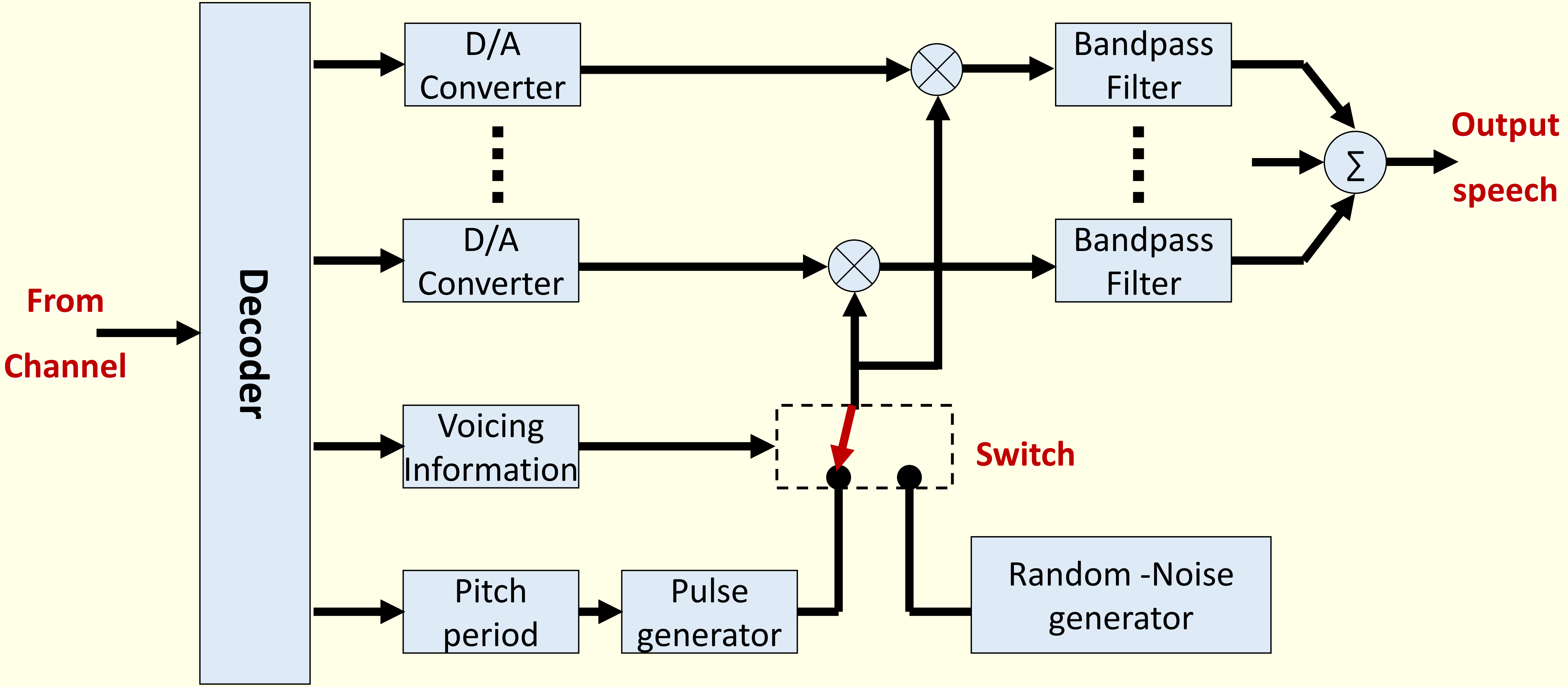
# The Channel Vocoder – analyzer (block diagram)



# The Channel Vocoder - synthesizer

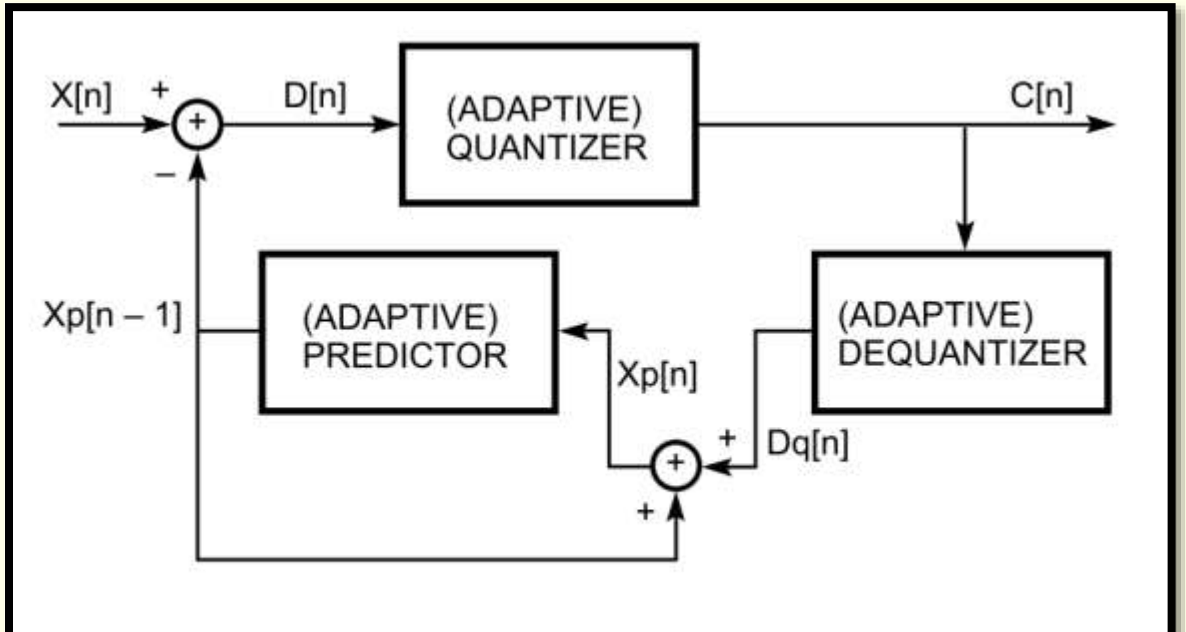
- At the receiver the signal samples are passed through D/A converters.
- The outputs of the D/As are multiplied by the voiced or unvoiced signal sources.
- The resulting signal are passed through bandpass filters.
- The outputs of the bandpass filters are summed to form the synthesized speech signal.

# The Channel Vocoder - synthesizer (block diagram)

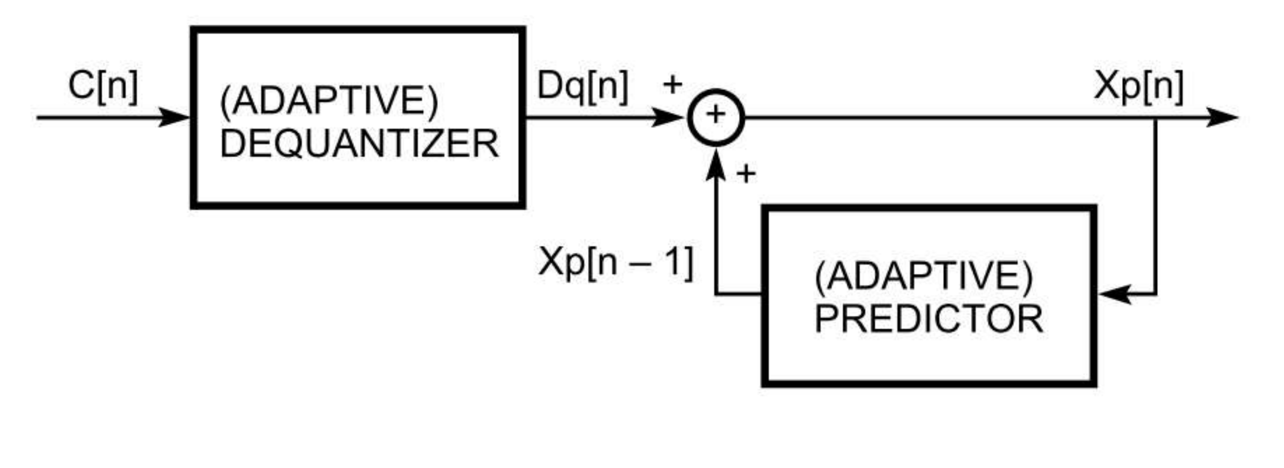


# Adaptive Differential Pulse Code Modulation

- The ADPCM coder takes advantage of the fact that neighboring audio samples are generally similar to each other. Instead of representing each audio sample independently as in Pulse Code Modulation (PCM), an ADPCM encoder computes the difference between each audio sample and its predicted value and outputs the PCM value of the differential.

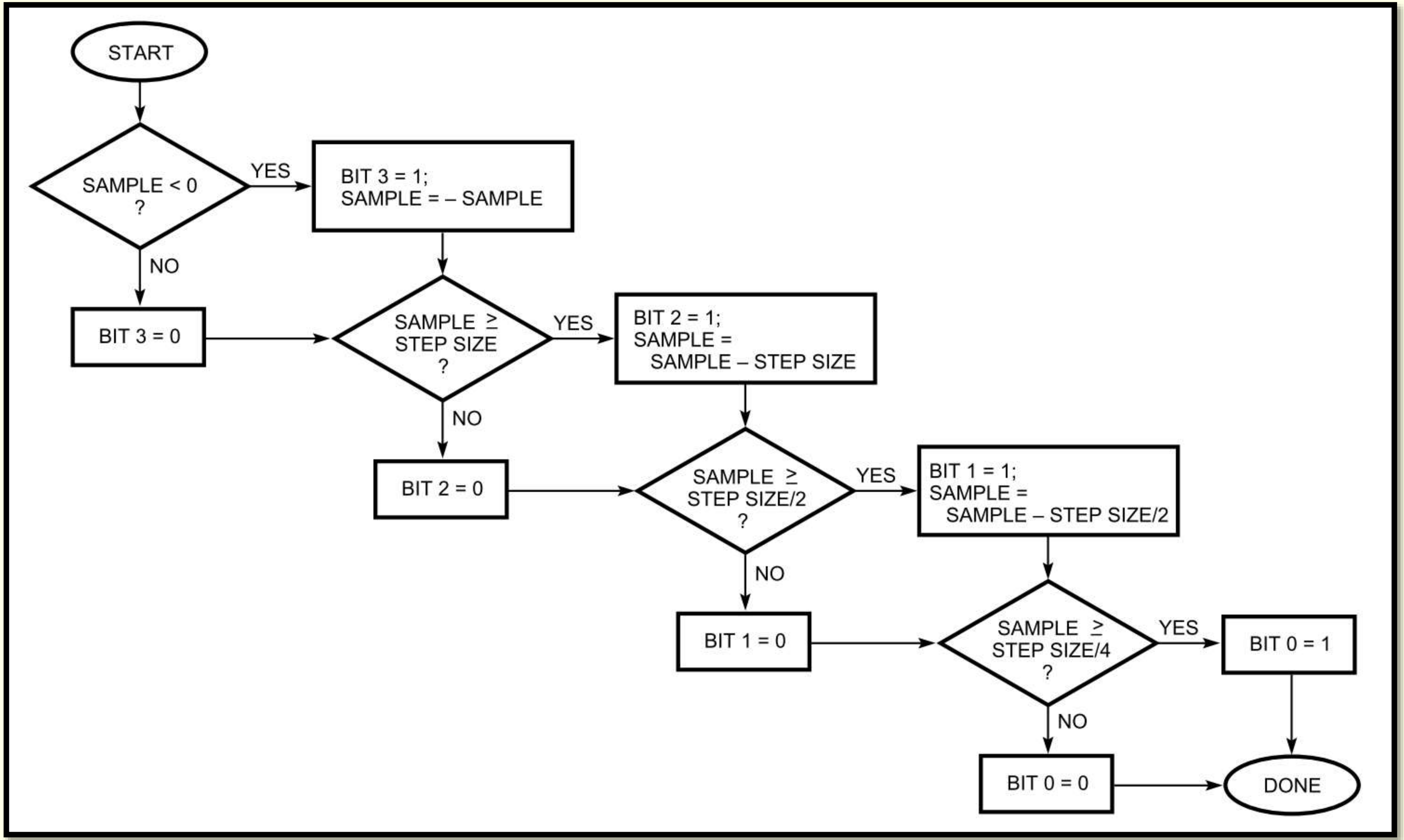


(a) ADPCM Encoder



(b) ADPCM Decoder

# ADPCM Quantization



# Linear Predictive Coding

- The objective of LP analysis is to estimate parameters of an all-pole model of the vocal tract.
- Several methods have been devised for generating the excitation sequence for speech synthesizes.
- LPC-type of speech analysis and synthesis are differ primarily in the type of excitation signal that is generated for speech synthesis.

# LPC 10

- This methods is called LPC-10 because of 10 coefficient are typically employed.
- LPC-10 partitions the speech into the 180 sample frame.
- Pitch and voicing decision are determined by using the **Average Magnitude Difference Function (**AMDF) and zero-crossing measures.

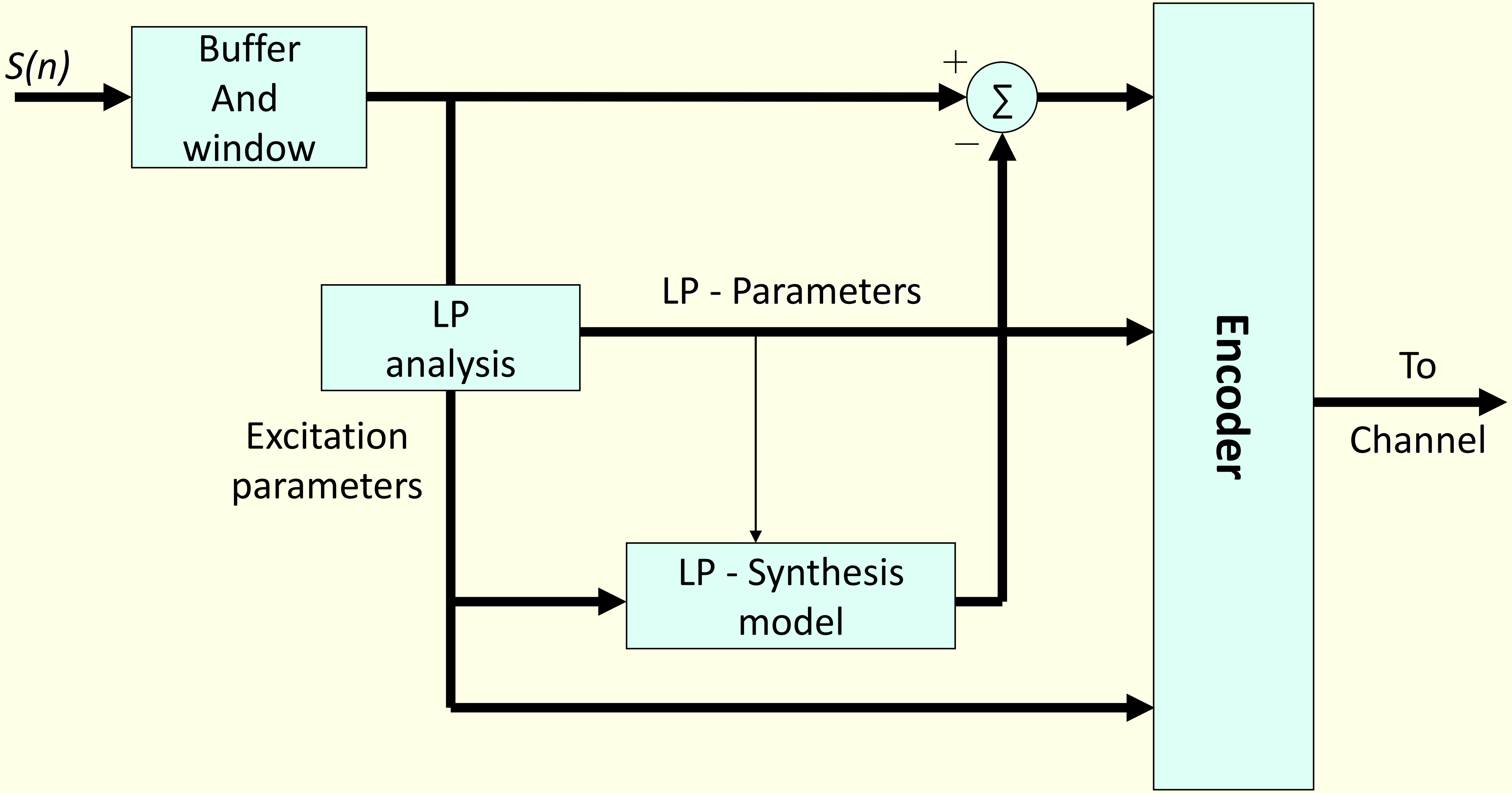
# Residual Excited LP Vocoder

- Speech quality in speech quality can be improved at the expense of a higher bit rate by computing and transmitting a residual error, as done in the case of DPCM.
- One method is that the LPC model and excitation parameters are estimated from a frame of speech.

# Residual Excited LP Vocoder

- The speech is synthesized at the transmitter and subtracted from the original speech signal to form the residual error.
- The residual error is quantized, coded, and transmitted to the receiver
- At the receiver the signal is synthesized by adding the residual error to the signal generated from the model.

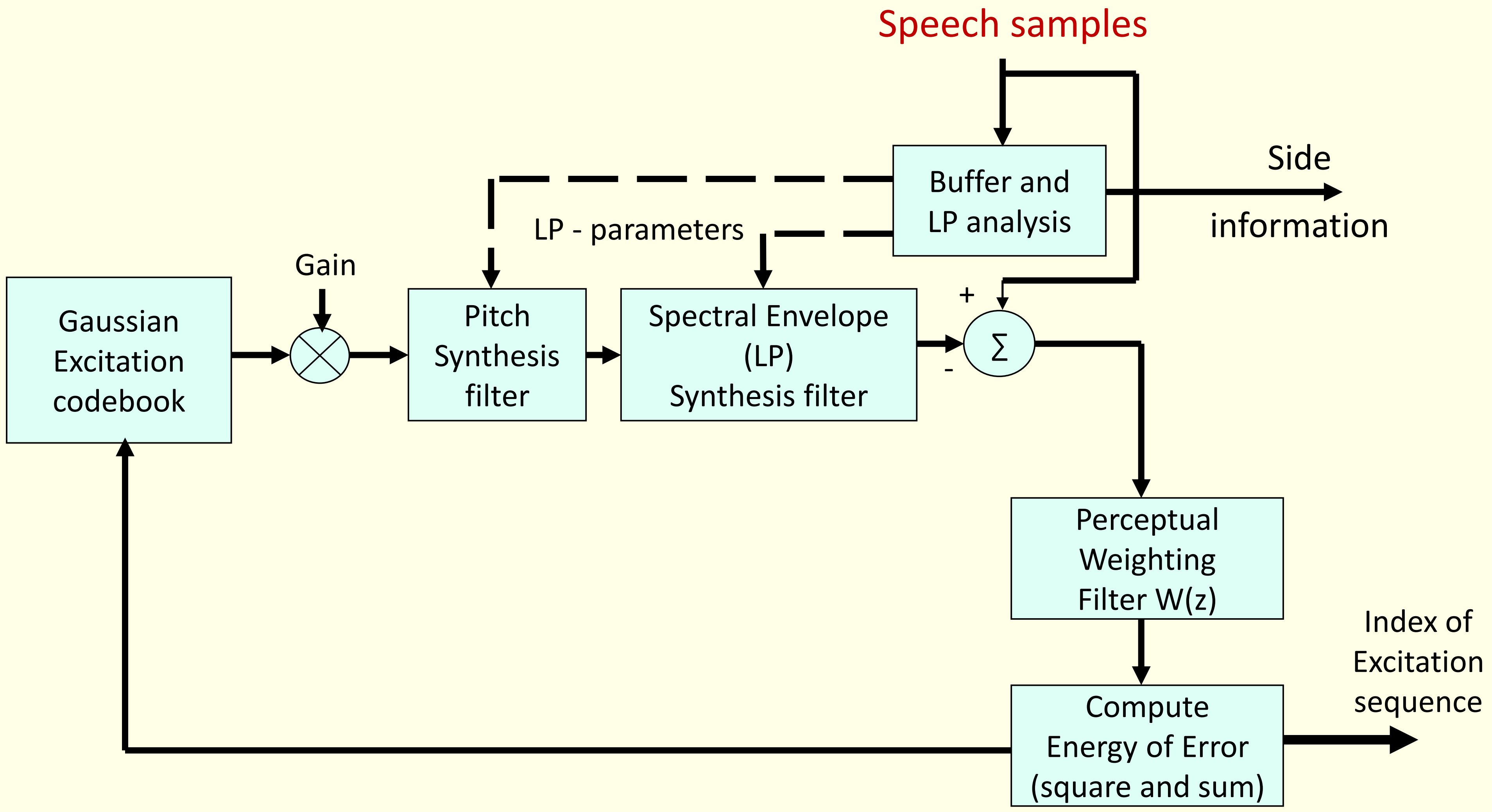
# RELP Block Diagram



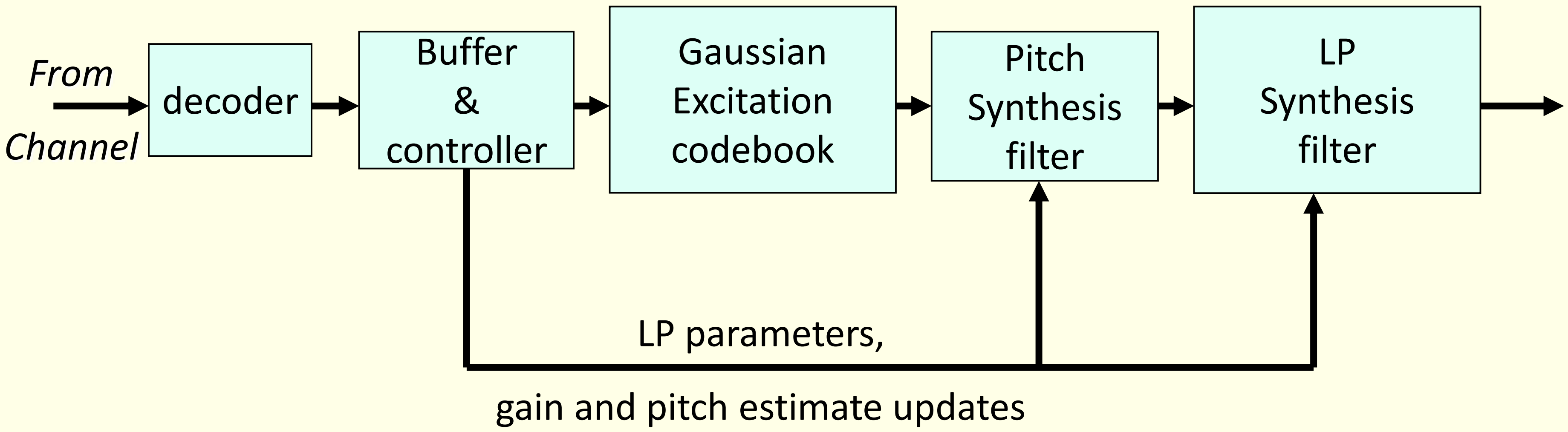
# Code Excited LP

- CELP is an analysis-by-synthesis method in which the excitation sequence is selected from a codebook of zero-mean Gaussian sequence.
- The bit rate of the CELP is 4800 bps.

# CELP (analysis-by-synthesis coder)

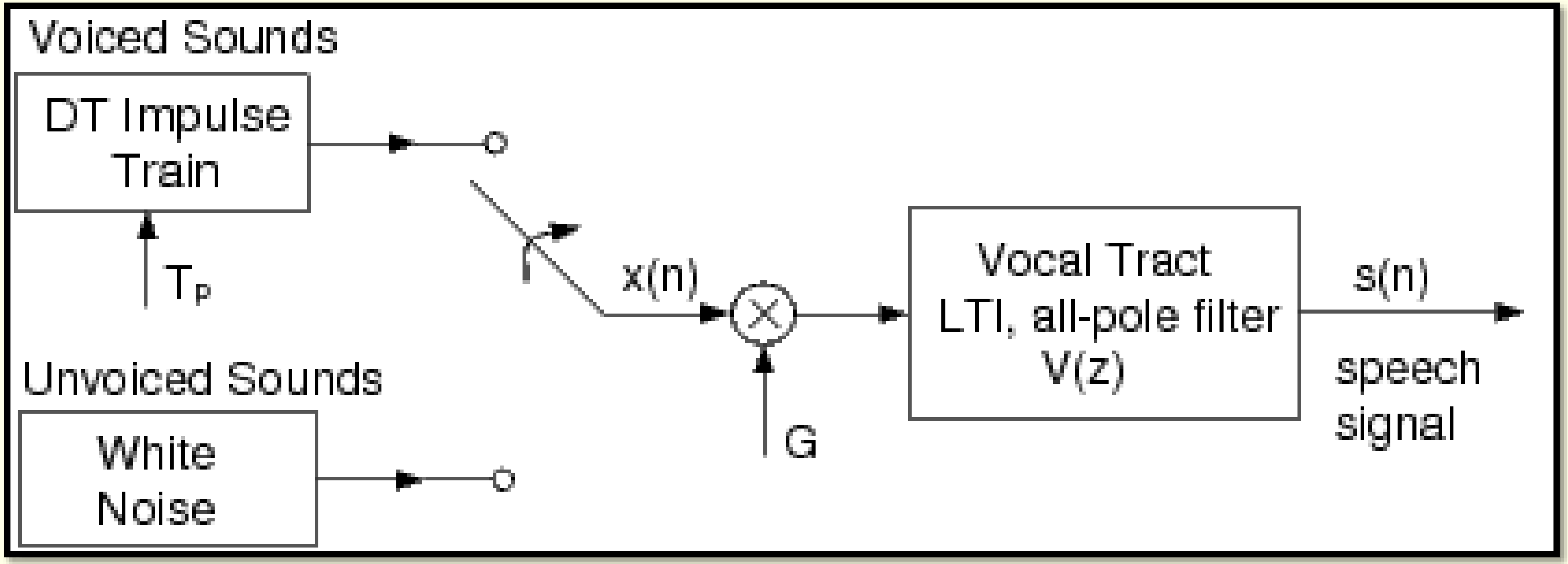


# CELP (synthesizer)

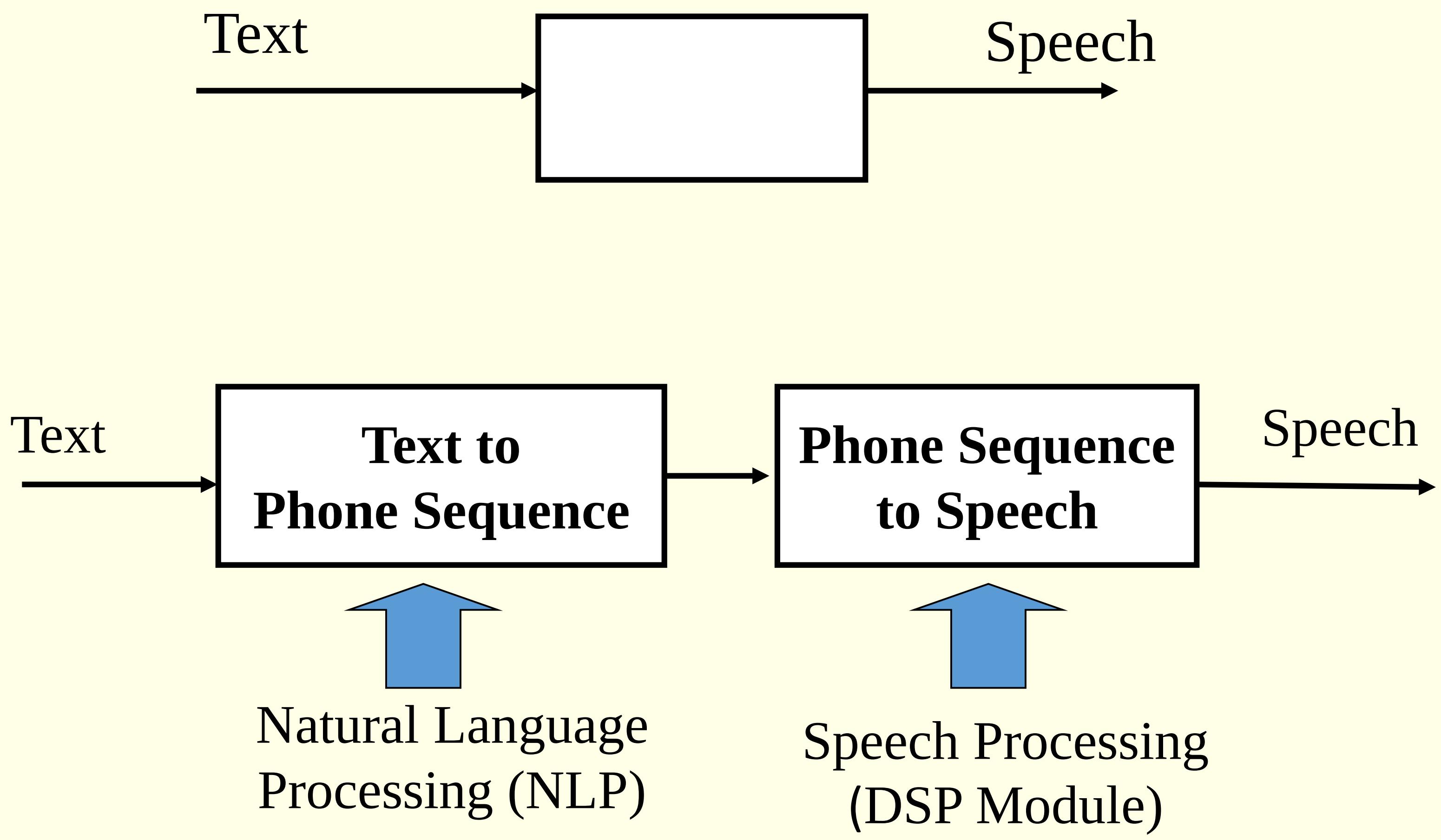


# 2. Text To Speech (TTS) Application

- A text to speech synthesizer is a computer based system that should be able to read any text whether it was directly introduced into the computer or through character recognition system (OCR). And speech should be intelligible and natural.



# Speech Synthesis Concept



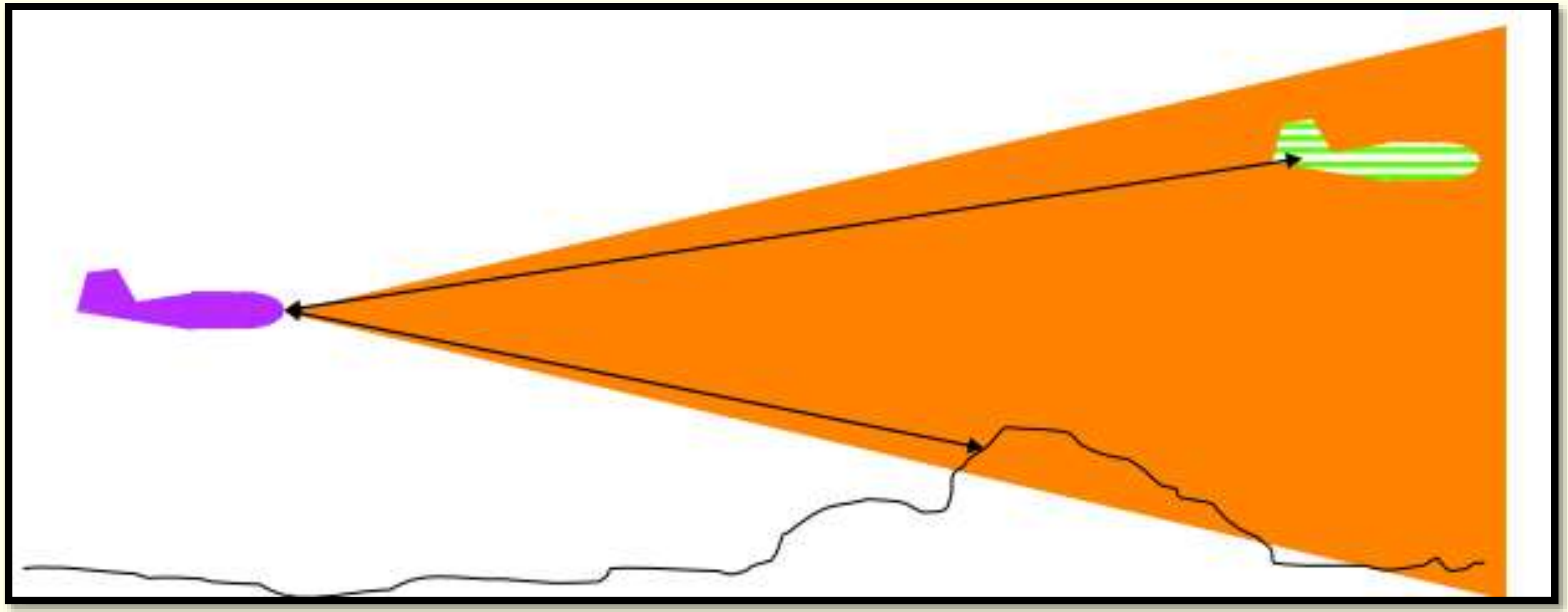
# NLP and DSP Modules

- The Natural Language Processing (NLP) module is capable of producing a phonetic transcription of the text to be read, together with the desired intonation and rhythm. It takes in the text as input and give narrow phonetic transcription as output which is further forwarded to the DSP module.
- the DSP module which transforms the symbolic information it receives into natural sounding speech. "Narrow phonetic transcription" which is taken as intermediate varies from synthesizer system to another.

# Radar Application

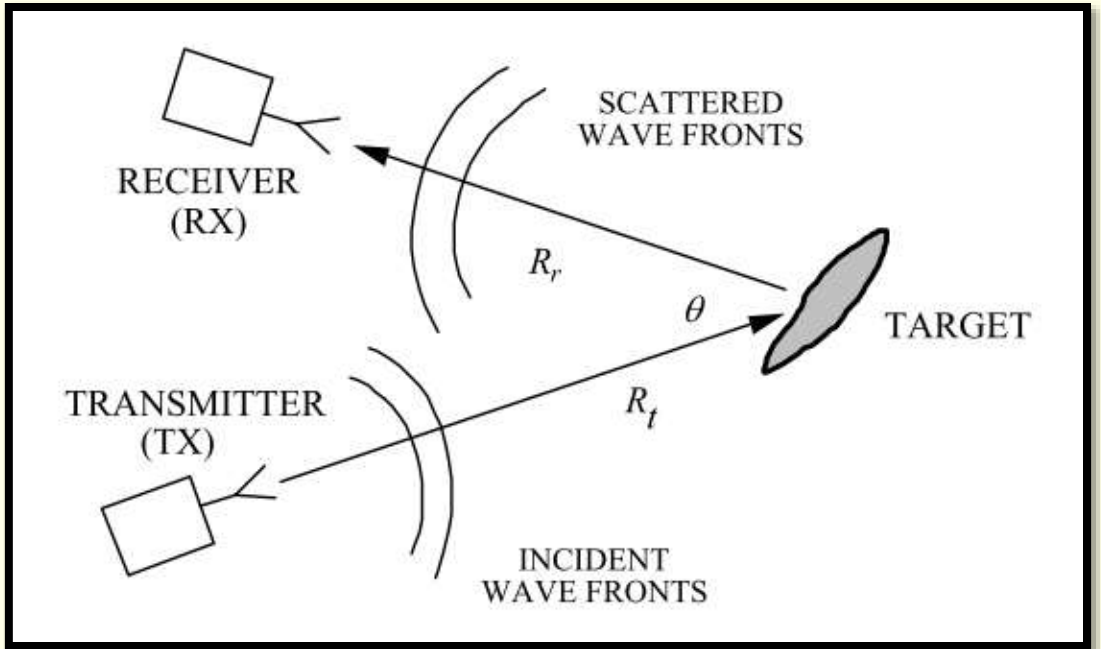
# Radar

- Radar: RAdio Detection And Ranging
- Need a directional radio beam.
- Measure time between transmit pulse and receive pulse
- Find Distance: Divide speed of light by interval time

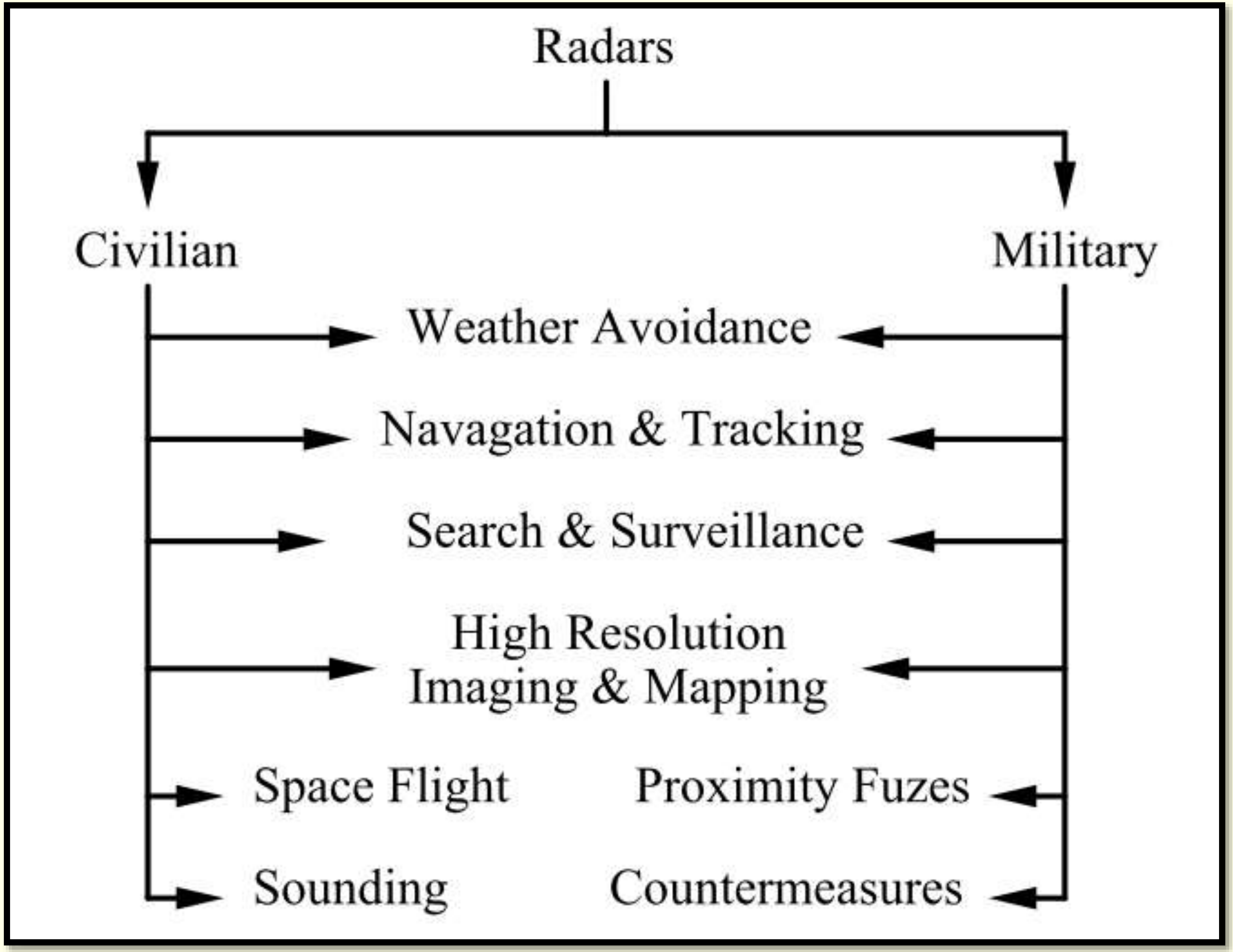


# Radio Detection and Ranging

- **<u>Bistatic</u>**: the transmit and receive antennas are at different locations as viewed from the target (e.g., ground transmitter and airborne receiver).
- **<u>Monostatic</u>**: the transmitter and receiver are colocated as viewed from the target (i.e., the same antenna is used to transmit and receive).
- **<u>Quasi-monostatic</u>**: the transmit and receive antennas are slightly separated but still appear to be at the same location as viewed from the target (e.g., separate transmit and receive antennas on the same aircraft).



# Classification by Function

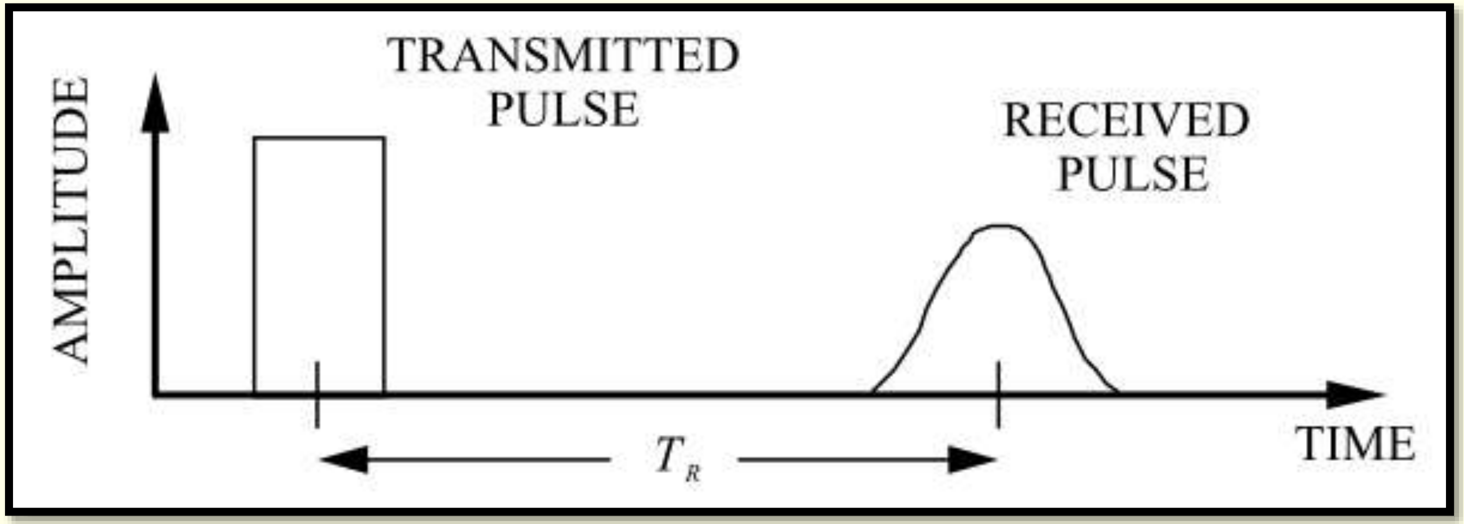


# Radar Functions

- Normal radar functions:
  1) range (from pulse delay)
  2) velocity (from Doppler frequency shift)
  3) angular direction (from antenna pointing)
- Signature analysis and inverse scattering:
  1) target size (from magnitude of return)
  2) target shape and components (return as a function of direction)
  3) moving parts (modulation of the return)
  4) material composition
- The complexity (cost & size) of the radar increases with the extent of the functions that the radar performs.

# Time Delay Ranging

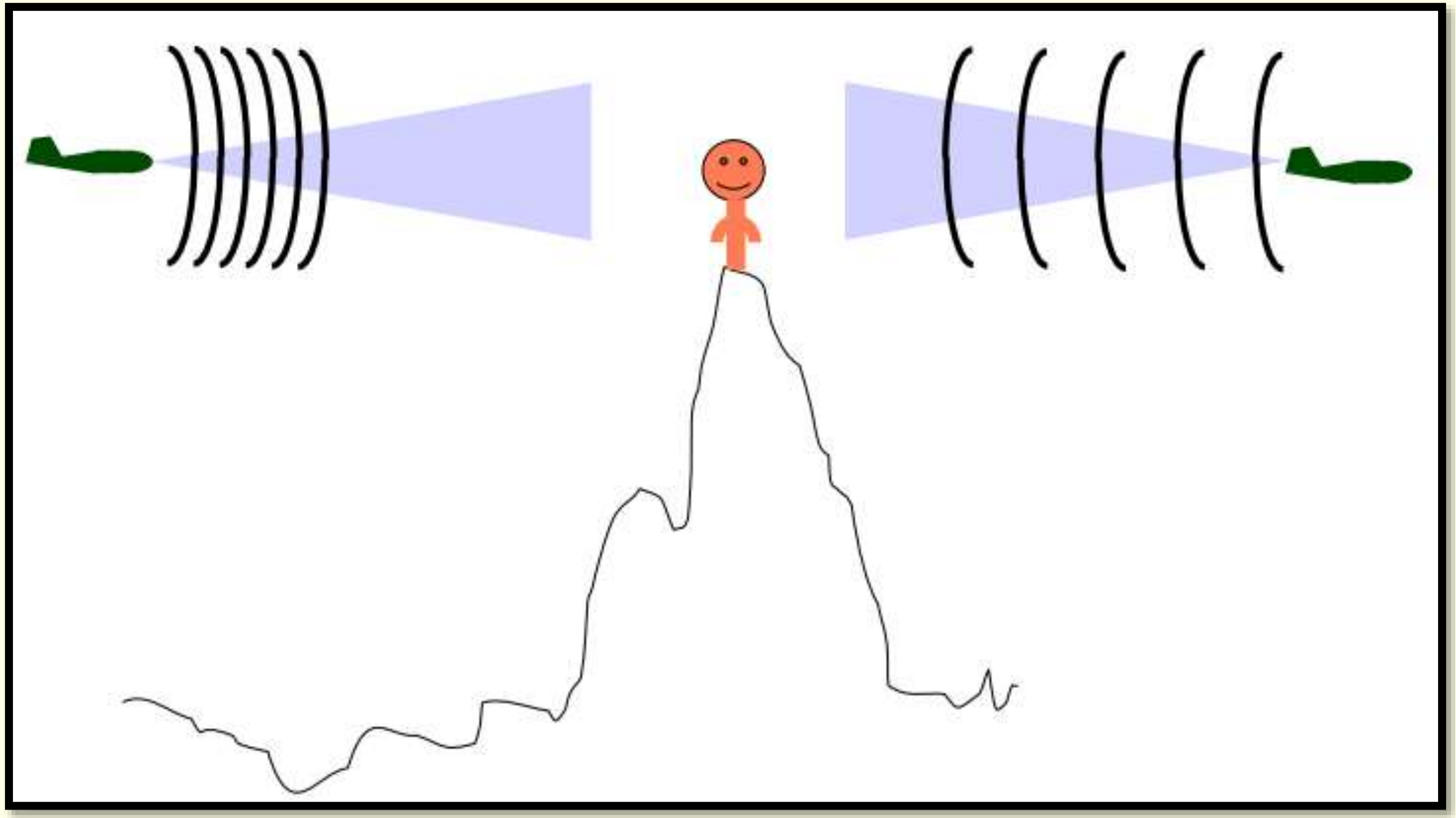
- Target range is the fundamental quantity measured by most radars. It is obtained by recording the round trip travel time of a pulse, $T_R$, and computing range from:
  - Bistatic: $R_t + R_r = c\,T_R$
  - Monostatic: $\mathrm{R} = \frac{c\,T_R}{2}$ $\quad (R_t = R_r = R)$
- where $c = 3 \times 10^8\, m/s$ is the velocity of light in free space.



# Doppler concept – frequency shift through motion

- The Doppler Effect is the change in the observed frequency of a source due to the relative motion between the source and the receiver.

- Frequency shift in received pulse: $f_{Doppler} = 2\ v_{relative}\ /\ \lambda$



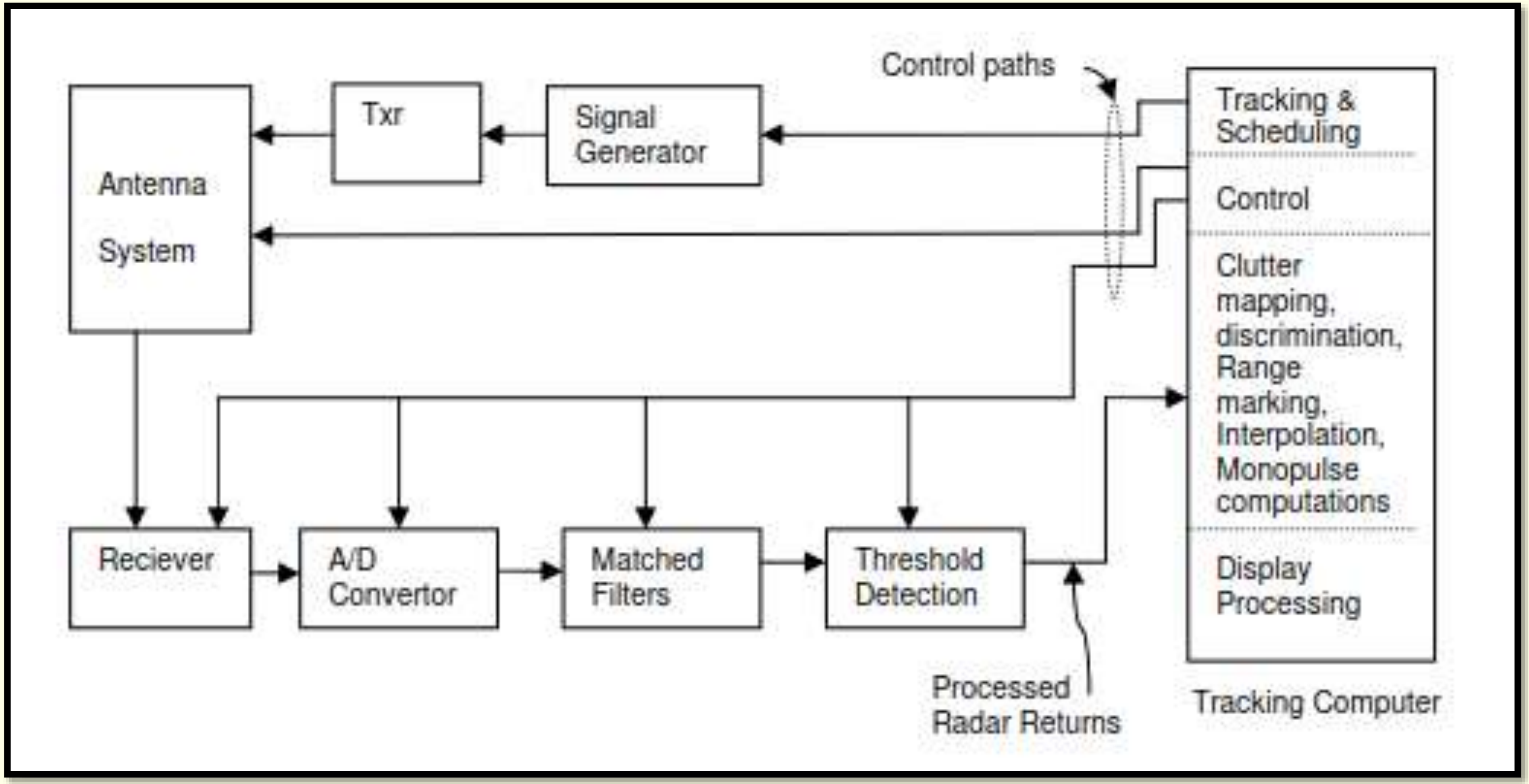
# Doppler effect

- Example: assume X band radar operating at 10GHz (3cm wavelength)
- Airborne radar traveling at 500 mph
  - Target 1 traveling away from radar at 800 mph
  - $V_{relative}$ = 500 – 800 = -300 mph = -134 meter/s

  - Target 2 traveling towards radar at 400 mph
  - $V_{relative}$ = 500 + 400 = 900 mph = 402 meter/s

  - First target Doppler shift = 2 (-134m/s) / (0.03m) = - 8.93 kHz
  - Second target Doppler shift = 2 (402m/s) / (0.03m) = 26.8 kHz

# Radar Block Diagram



# Wiener Filter

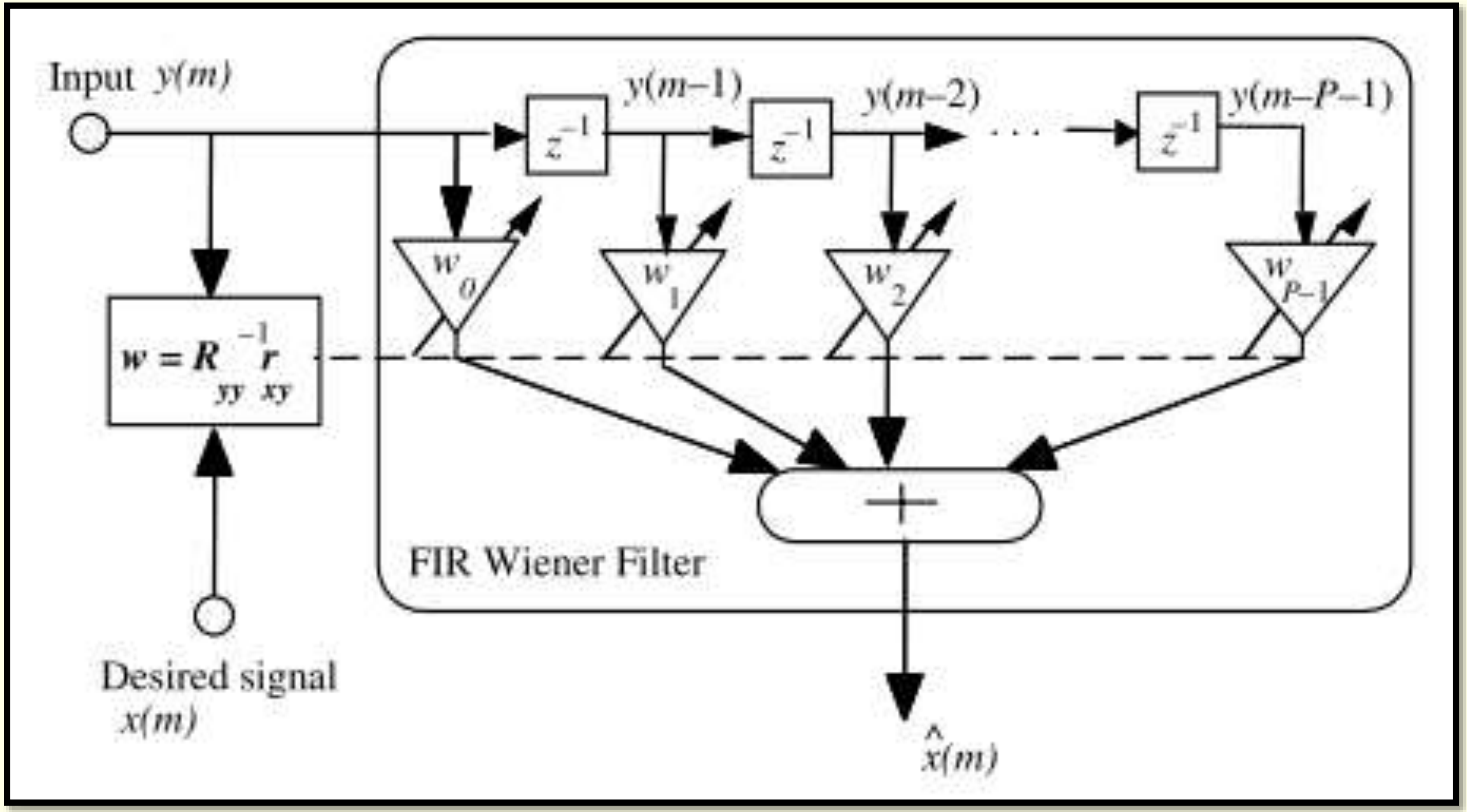
# The Wiener Filter

- The Wiener filter solves the signal estimation problem for stationary signals.
- The filter was introduced by Norbert Wiener in the 1940's. A major contribution was the use of a statistical model for the estimated signal (the Bayesian approach!).
- Wiener filters play a central role in a wide range of applications such as linear prediction, echo cancellation, signal restoration, channel equalization and system identification.
- The coefficients of a Wiener filter are calculated to minimize the average squared distance between the filter output and a desired signal.
- The filter is optimal in the sense of the Minimum Mean Square Error (MMSE).
- As we shall see, the Kalman filter solves the corresponding filtering problem in greater generality, for non-stationary signals.
- We shall focus here on the discrete-time version of the Wiener filter.

# Wiener Filters: Least Square Error Estimation

- A Wiener filter can be
  - IIR-filter
    - The formulation of an IIR Wiener filter results in a set of non-linear equations
  - FIR-filter
    - The formulation of an FIR Wiener filter results in a set of linear equations and has a closed-form solution.
- In this chapter, we consider FIR Wiener filters, since they are relatively simple to compute, inherently stable and more practical.
- The main drawback of FIR filters compared with IIR filters is that they may need a large number of coefficients to approximate a desired response.

# Wiener filter structure.



# Wiener filter

- The Wiener filter error signal, e(m) is defined as the difference between the desired signal x(m) and the filter output signal

$$\begin{aligned} e(m) &= x(m) - \hat{x}(m) \\ &= x(m) - \mathbf{w}^{\mathrm{T}} \mathbf{y} \end{aligned} \qquad (6.2)$$

- In Equation (6.2), for a given input signal y(m) and a desired signal x(m), the filter error e(m) depends on the filter coefficient vector w.

$$\begin{pmatrix} e(0) \\ e(1) \\ e(2) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(N-1) \end{pmatrix} - \begin{pmatrix} y(0) & y(-1) & y(-2) & \cdots & y(1-P) \\ y(1) & y(0) & y(-1) & \cdots & y(2-P) \\ y(2) & y(1) & y(0) & \cdots & y(3-P) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & y(N-3) & \cdots & y(N-P) \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{P-1} \end{pmatrix} \qquad (6.3)$$

# Wiener filter

- In a compact vector notation this matrix equation may be written as

$$e = x - Yw \tag{6.4}$$

- where
  - e is the error vector, x is the desired signal vector, and Y is the input signal matrix.
  - It is assumed that the P initial input signal samples $[y(-1), \ldots, y(-P-1)]$ are either known or set to zero.
- In Equation (6.3), if the number of signal samples is
  - $N < P$, the matrix equation is said to be underdetermined.
  - $N = P$, the matrix equation has a unique filter solution .
  - $N > P$, the matrix equation is said to be overdetermined.

# Wiener filter

- The Wiener filter coefficients are obtained by minimizing an average squared error function $E\{e^2(m)\}$ with respect to the filter coefficient vector w. From Equation (6.2), the mean square estimation error is given by

$$\begin{aligned} \mathcal{E}[e^2(m)] &= \mathcal{E}[(x(m) - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{y})^2] \\ &= \mathcal{E}[x^2(m)] - 2\boldsymbol{w}^{\mathrm{T}}\mathcal{E}[\boldsymbol{y}x(m)] + \boldsymbol{w}^{\mathrm{T}}\mathcal{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}]\boldsymbol{w} \\ &= r_{xx}(0) - 2\boldsymbol{w}^{\mathrm{T}}\boldsymbol{r}_{yx} + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{yy}\boldsymbol{w} \end{aligned} \qquad (6.5)$$

  - where $R_{yy} = E[y(m)y^T(m)]$ is the autocorrelation matrix of the input signal and $r_{xy} = E[x(m)\, y(m)]$ is the cross-correlation vector of the input and the desired signals.

- An expanded form of Equation (6.5) can be obtained as

$$\mathcal{E}[e^2(m)] = r_{xx}(0) - 2\sum_{k=0}^{P-1} w_k r_{yx}(k) + \sum_{k=0}^{P-1} w_k \sum_{j=0}^{P-1} w_j r_{yy}(k-j) \qquad (6.6)$$

# Wiener filter

- From Equation (6.5), the mean square error for an FIR filter is a quadratic function of the filter coefficient vector w and has a single minimum point.

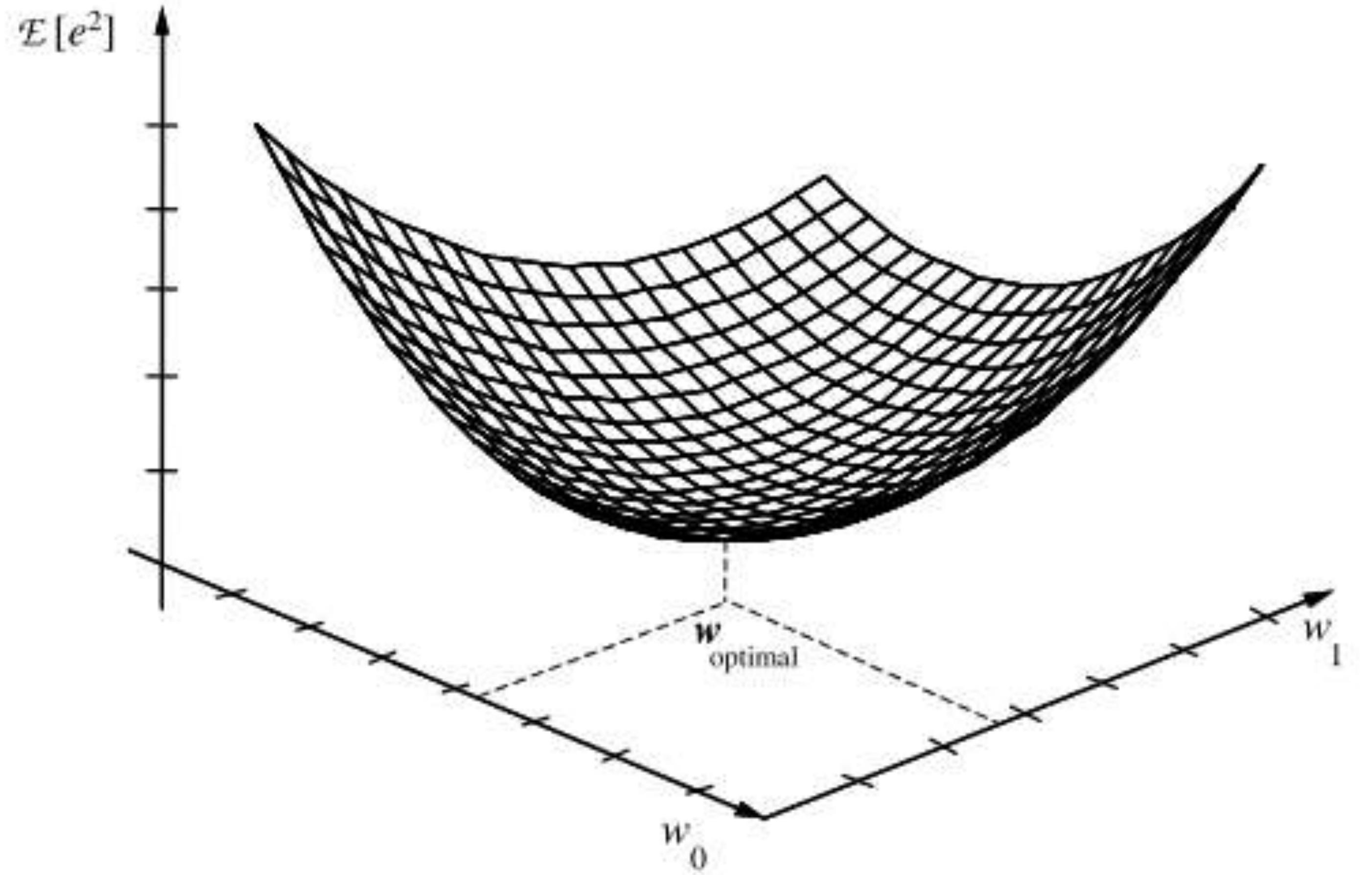



**Figure 6.2** Mean square error surface for a two-tap FIR filter.

# Wiener filter

- At this optimal operating point the mean square error surface has zero gradient. From Equation (6.5), the gradient of the mean square error function with respect to the filter coefficient vector is given by

$$\frac{\partial}{\partial \boldsymbol{w}} \mathcal{E}[e^2(m)] = -2\mathcal{E}[x(m)\boldsymbol{y}(m)] + 2\boldsymbol{w}^{\mathrm{T}}\mathcal{E}[\boldsymbol{y}(m)\boldsymbol{y}^{\mathrm{T}}(m)]$$
$$= -2\boldsymbol{r}_{yx} + 2\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{yy} \tag{6.7}$$

- where the gradient vector is defined as

$$\frac{\partial}{\partial \boldsymbol{w}} = \left[\frac{\partial}{\partial w_0}, \frac{\partial}{\partial w_1}, \frac{\partial}{\partial w_2}, \ldots, \frac{\partial}{\partial w_{P-1}}\right]^{\mathrm{T}} \tag{6.8}$$

- The minimum mean square error Wiener filter is obtained by setting Equation (6.7) to zero:

$$\boldsymbol{R}_{yy}\,\boldsymbol{w} = \boldsymbol{r}_{yx} \tag{6.9}$$

- or, equivalently,

$$\boldsymbol{w} = \boldsymbol{R}_{yy}^{-1}\,\boldsymbol{r}_{yx} \tag{6.10}$$

# Wiener filter

- In an expanded form, the Wiener filter solution Equation (6.10) can be written as

$$\begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{P-1} \end{pmatrix} = \begin{pmatrix} r_{yy}(0) & r_{yy}(1) & r_{yy}(2) & \cdots & r_{yy}(P-1) \\ r_{yy}(1) & r_{yy}(0) & r_{yy}(1) & \cdots & r_{yy}(P-2) \\ r_{yy}(2) & r_{yy}(1) & r_{yy}(0) & \cdots & r_{yy}(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{yy}(P-1) & r_{yy}(P-2) & r_{yy}(P-3) & \cdots & r_{yy}(0) \end{pmatrix}^{-1} \begin{pmatrix} r_{yx}(0) \\ r_{yx}(1) \\ r_{yx}(2) \\ \vdots \\ r_{yx}(P-1) \end{pmatrix} \quad (6.11)$$

- From Equation (6.11), the calculation of the Wiener filter coefficients requires the <u>autocorrelation matrix of the input signal</u> and <u>the cross-correlation vector of the input and the desired signals</u>.

# Wiener filter

- For a signal record of length N samples, the time-averaged correlation values are computed as

$$r_{yy}(k)=\frac{1}{N}\sum_{m=0}^{N-1} y(m)\,y(m+k) \qquad (6.12)$$

# 3- Speech enhancement

- Wiener Filtering:
  - A linear estimation of clean signal from the noisy signal Using MMSE criterion

| | | |
|---|---|---|
| $y_t$ | Clean Speech | $\mathrm{K} \times 1$ |
| $v_t$ | Noise | $\mathrm{K} \times 1$ |
| $z_t$ | Noisy Speech | $\mathrm{K} \times 1$ |

$$z_t = y_t + v_t \qquad \text{For additive noise}$$

$$\hat{y}_t = E\{y_t | z_t\} = a z_t = a(y_t + v_t)$$

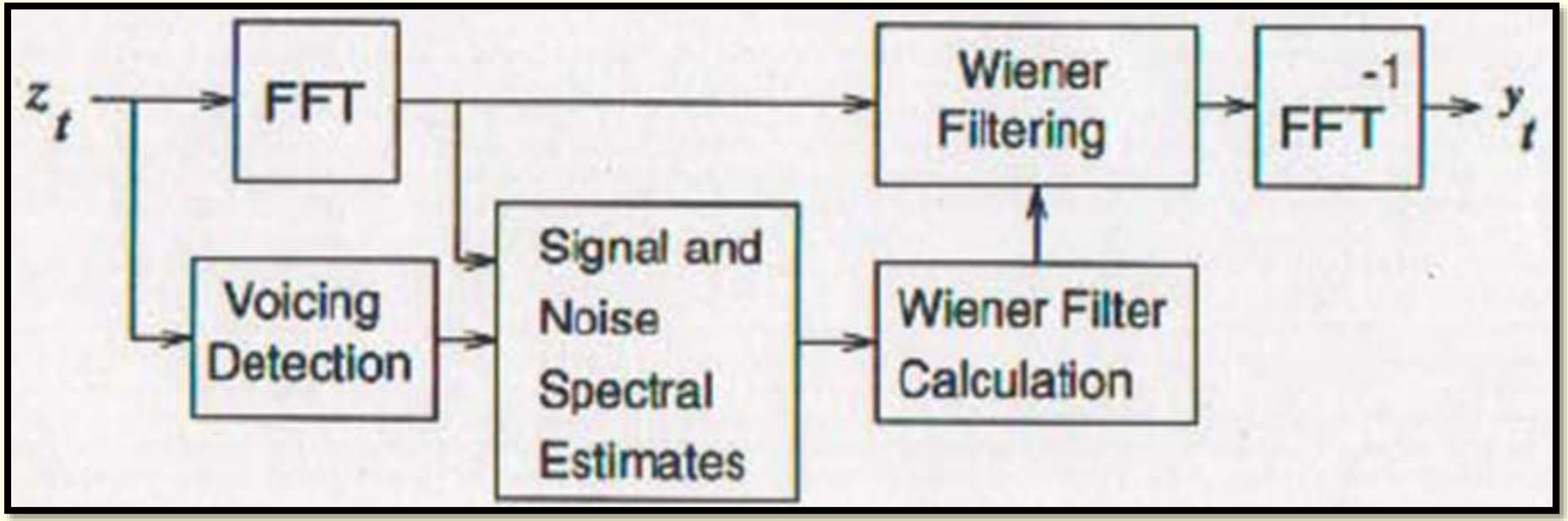
# Wiener Filter

$$S_{zz}(\omega) = S_{yy}(\omega) + S_{vv}(\omega)$$

$$H(j\omega) = \frac{S_{yy}(\omega)}{S_{yy}(\omega) + S_{vv}(\omega)}$$

$$H(j\omega) = \frac{S_{zz}(\omega) - S_{vv}(\omega)}{S_{zz}(\omega)}$$

# Spectral Subtraction Method



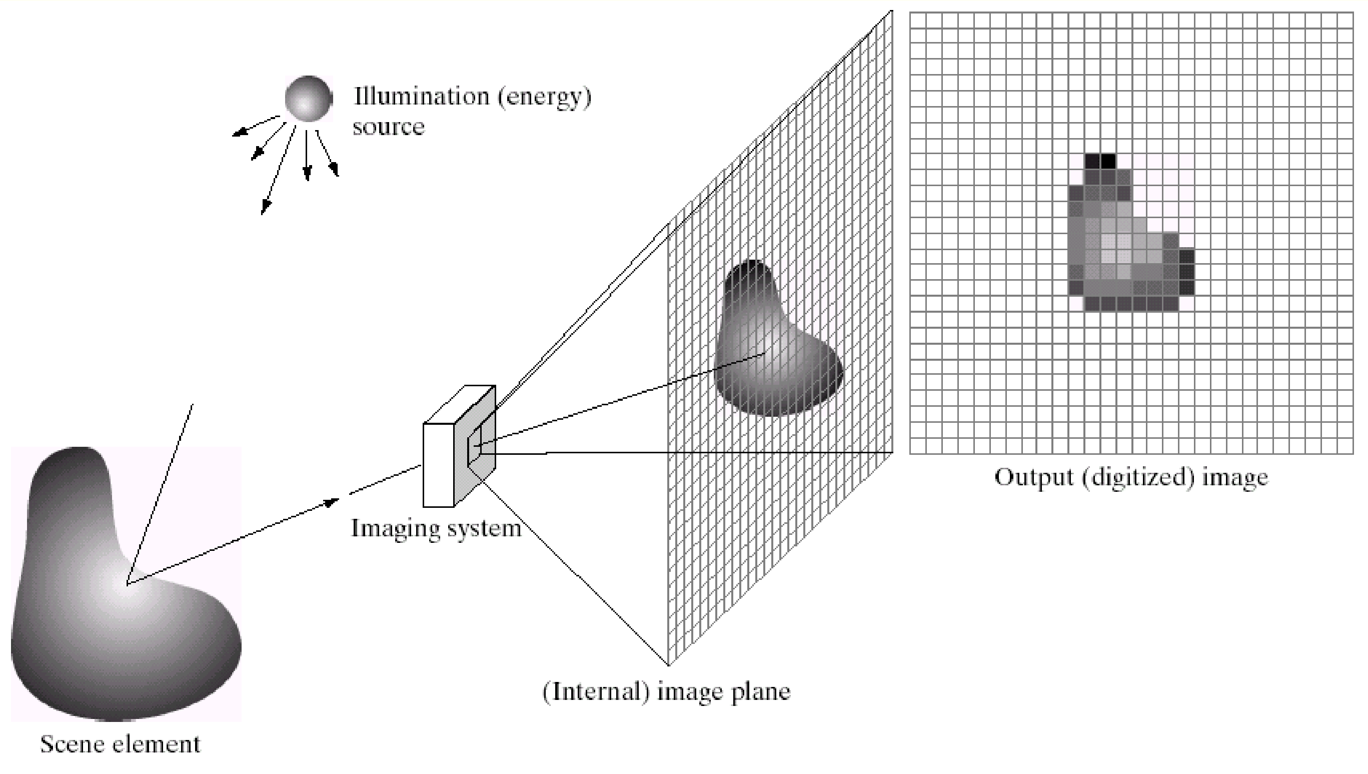
# Introduction 2 Image Processing

# Agenda

- Introduction 2 Image Processing
  - What is a digital image?
  - What is digital image processing?
  - Examples of DIP
  - Key stages in DIP
- Image Compression
  - Elementary Information Theory
  - Lossless Compression
  - JPEG Compression

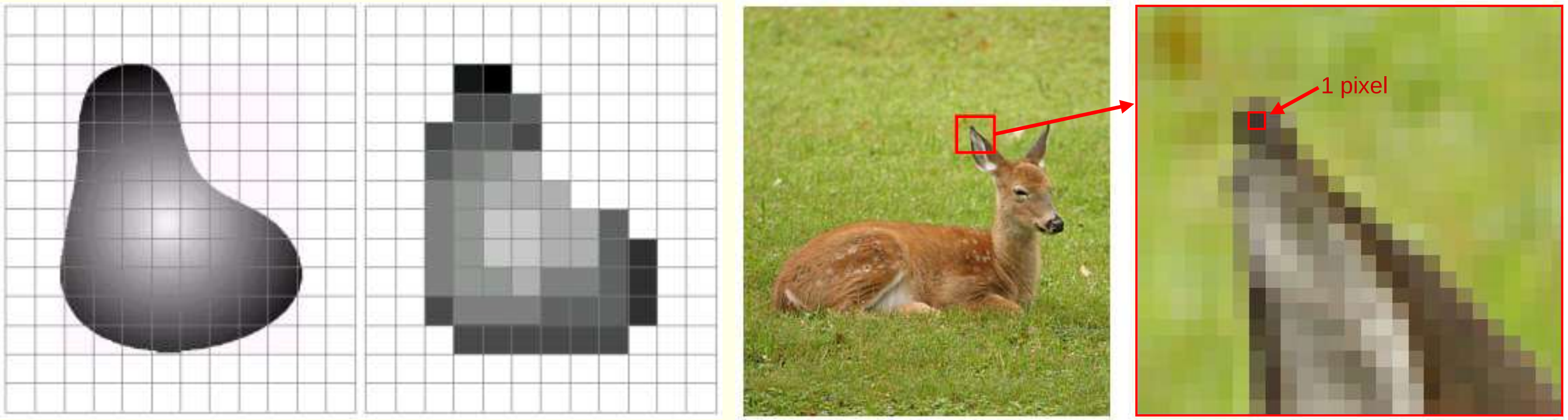
# What is a Digital Image?

- A digital image is a representation of a two-dimensional image as a finite set of digital values, called picture elements or pixels



# What is a Digital Image? (cont…)

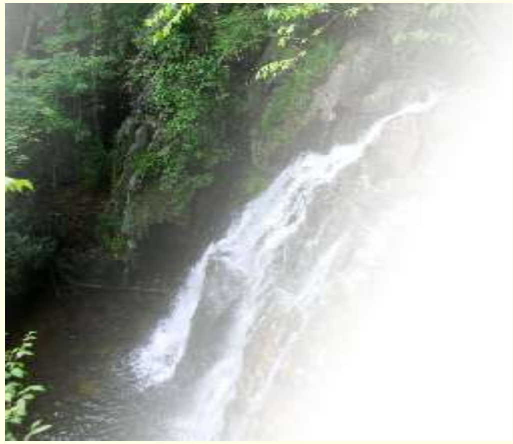
- Pixel values typically represent gray levels, colors, heights, opacities etc.
- Remember digitization implies that a digital image is an approximation of a real scene



# What is a Digital Image? (cont…)

- Common image formats include:
  - 1 sample per point (B&W or Grayscale)
  - 3 samples per point (Red, Green, and Blue)
  - 4 samples per point (Red, Green, Blue, and “Alpha”, a.k.a. Opacity)





# What is Digital Image Processing?

- Digital image processing focuses on two major tasks
  - Improvement of pictorial information for human interpretation
  - Processing of image data for storage, transmission and representation for autonomous machine perception
- Some argument about where image processing ends and fields such as image analysis and computer vision start

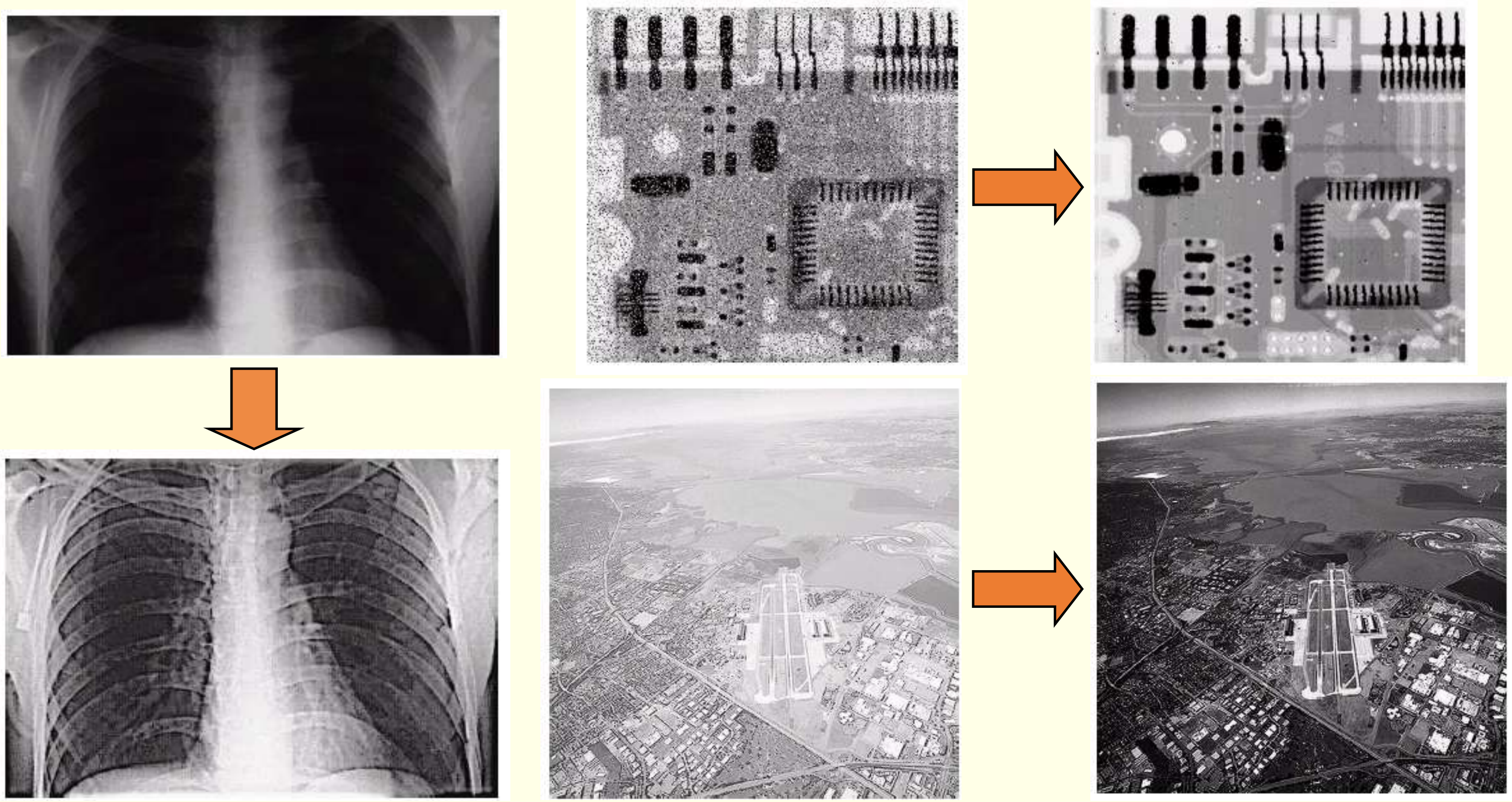
# What is DIP? (cont…)

- The continuum from image processing to computer vision can be broken up into low-, mid- and high-level processes

| **Low Level Process** | **Mid Level Process** | **High Level Process** |
|---|---|---|
| **Input:** Image<br>**Output:** Image | **Input:** Image<br>**Output:** Attributes | **Input:** Attributes<br>**Output:** Understanding |
| **Examples:** Noise removal, image sharpening | **Examples:** Object recognition, segmentation | **Examples:** Scene understanding, autonomous navigation |

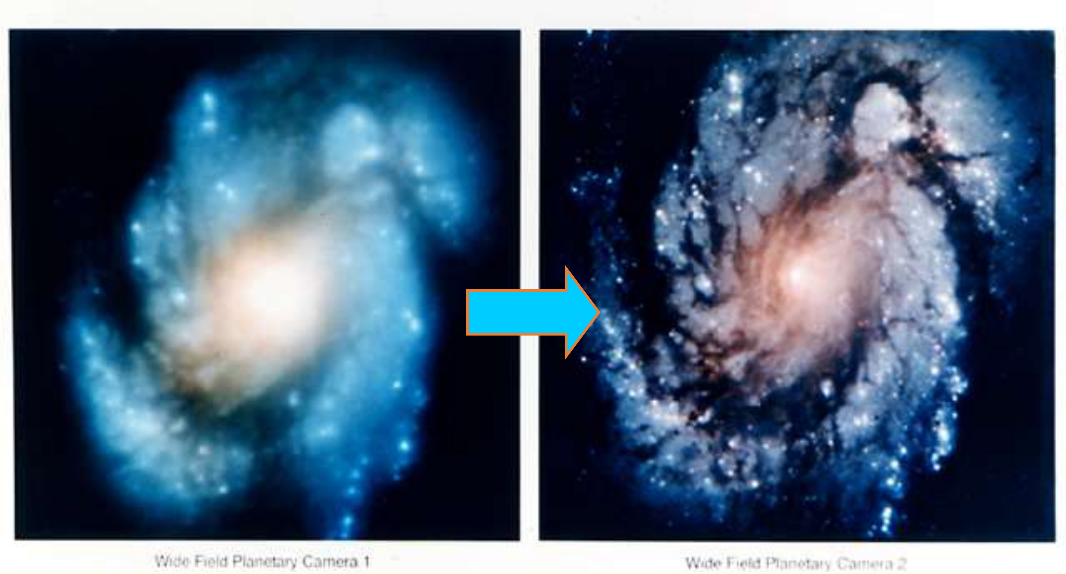
# Examples: Image Enhancement

- One of the most common uses of DIP techniques: improve quality, remove noise etc.



# Examples: The Hubble Telescope

- Launched in 1990 the Hubble telescope can take images of very distant objects
- However, an incorrect mirror made many of Hubble's images useless
- Image processing techniques were used to fix this

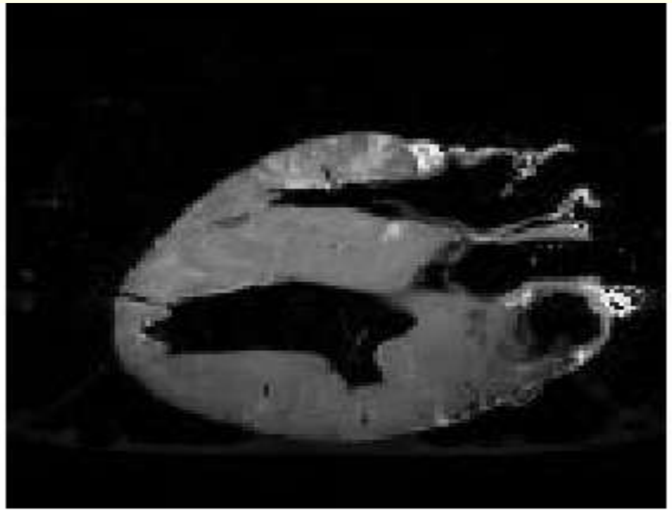


# Examples: Artistic Effects

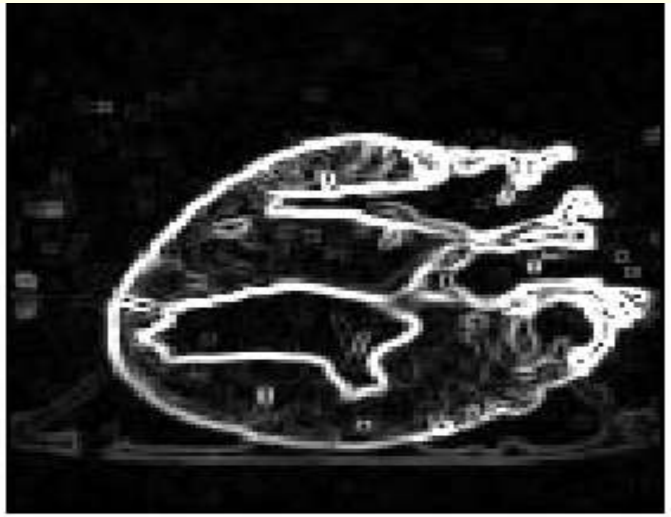
- Artistic effects are used to make images more visually appealing, to add special effects and to make composite images

# Examples: Medicine

- Take slice from MRI scan of canine heart, and find boundaries between types of tissue
  - Image with gray levels representing tissue density
  - Use a suitable filter to highlight edges



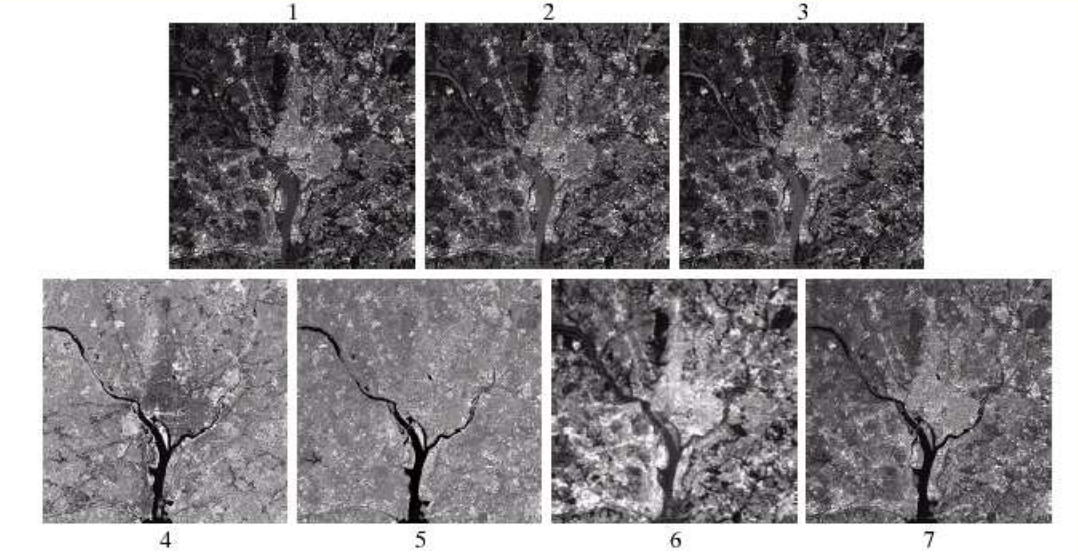
Original MRI Image of a Dog Heart



Edge Detection Image

# Examples: GIS

- Geographic Information Systems
  - Digital image processing techniques are used extensively to manipulate satellite imagery
  - Terrain classification
  - Meteorology

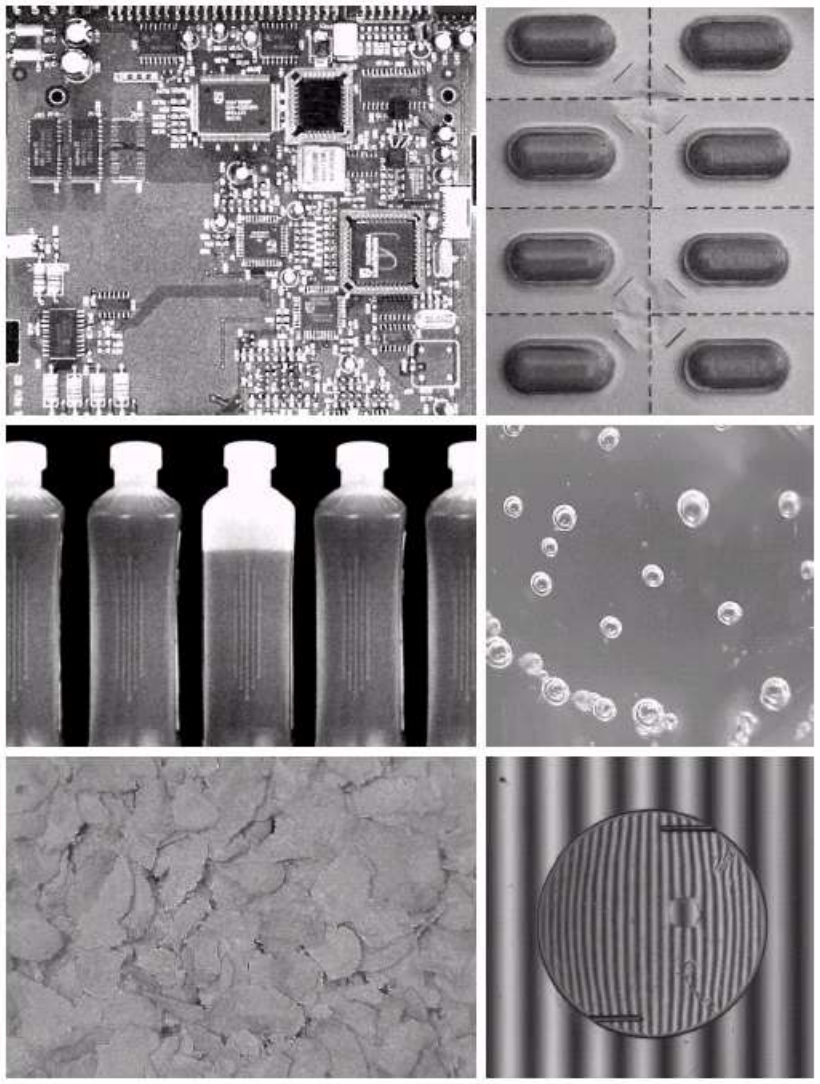



# Examples: GIS (cont…)

- Night-Time Lights of the World data set
  - Global inventory of human settlement
  - Not hard to imagine the kind of analysis that might be done using this data

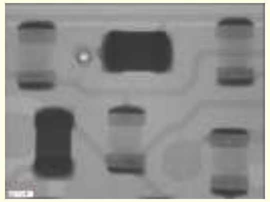
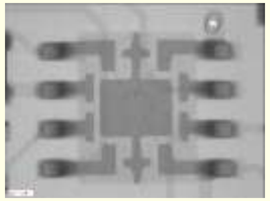
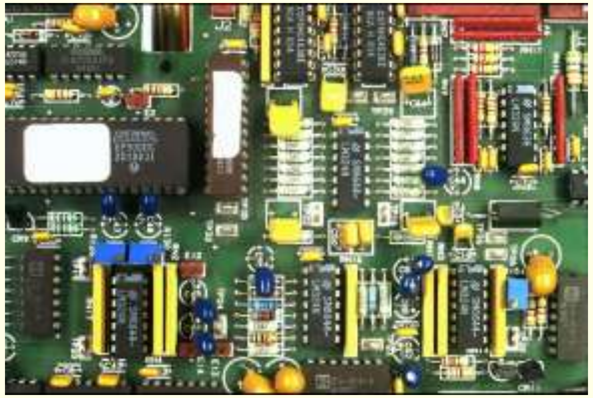
# Examples: Industrial Inspection

- Human operators are expensive, slow and unreliable
- Make machines do the job instead
- Industrial vision systems are used in all kinds of industries
- Can we trust them?



# Examples: PCB Inspection

- Printed Circuit Board (PCB) inspection
  - Machine inspection is used to determine that all components are present and that all solder joints are acceptable
  - Both conventional imaging and x-ray imaging are used

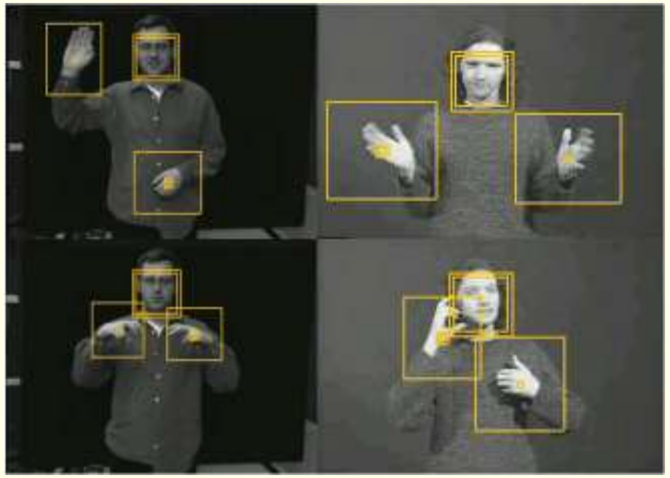






# Examples: Law Enforcement

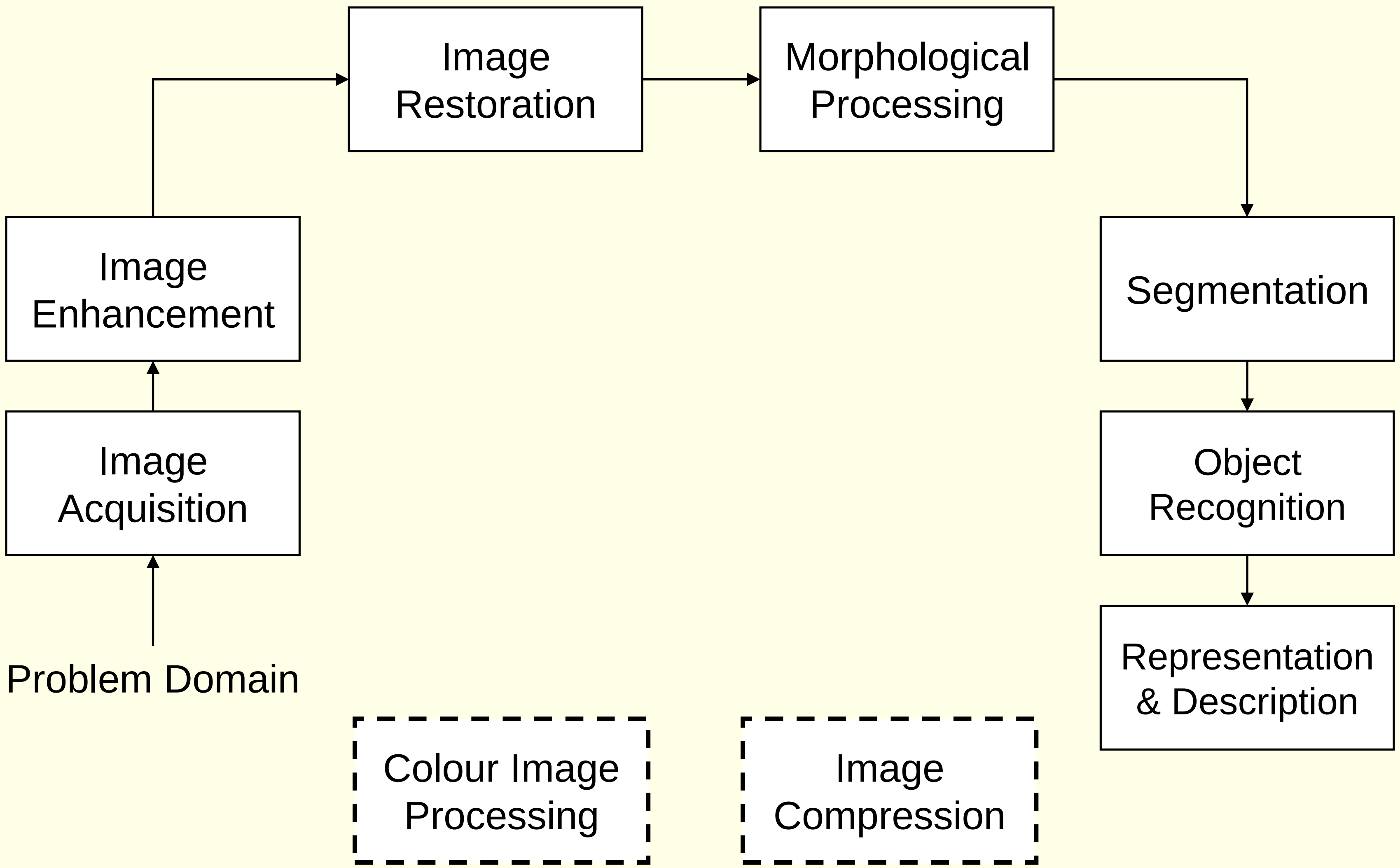
- Image processing techniques are used extensively by law enforcers
  - Number plate recognition for speed cameras/automated toll systems
  - Fingerprint recognition
  - Enhancement of CCTV images


# Examples: HCI

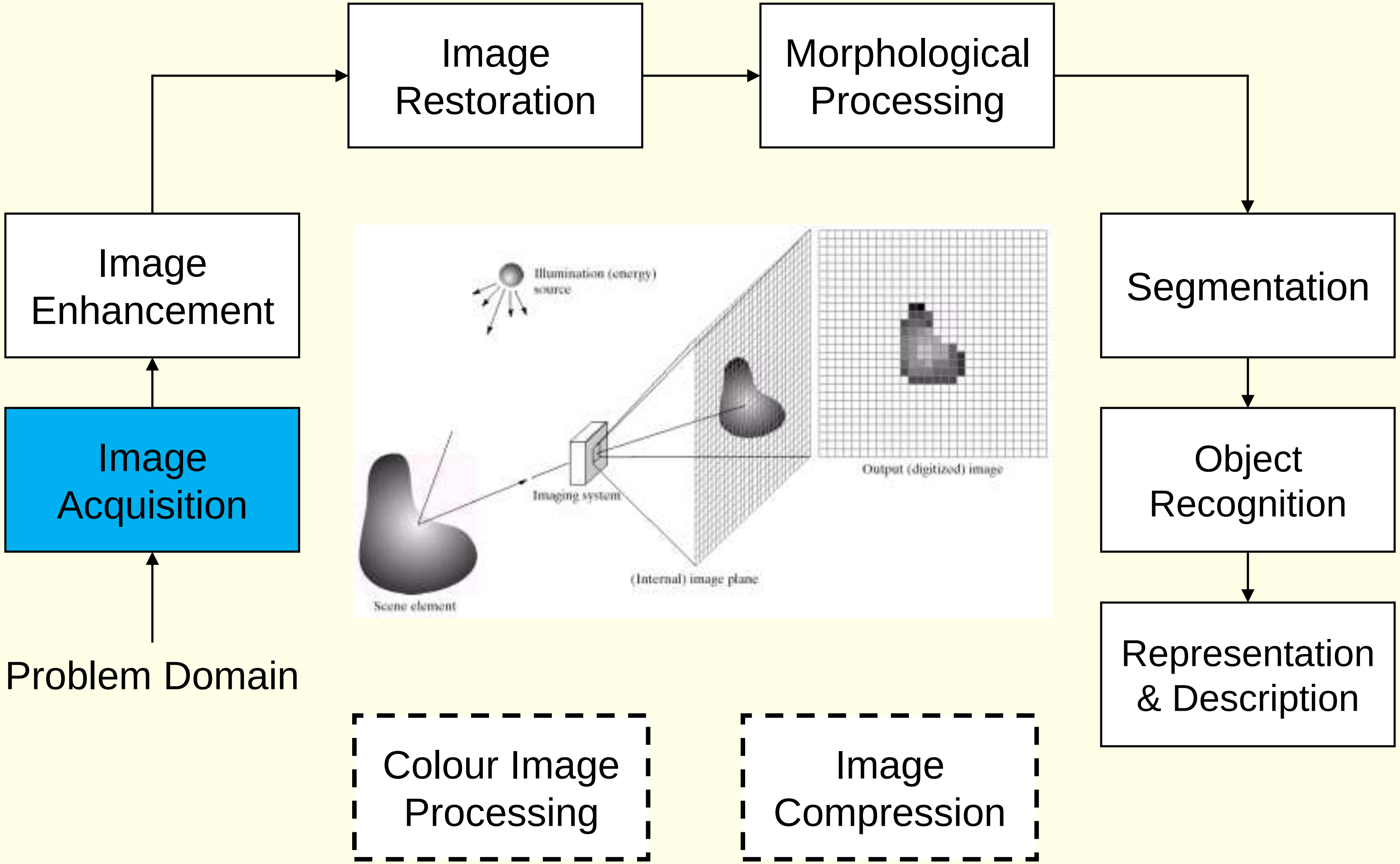
- Try to make human computer interfaces more natural
  - Face recognition
  - Gesture recognition
- Does anyone remember the user interface from "Minority Report"?
- These tasks can be extremely difficult



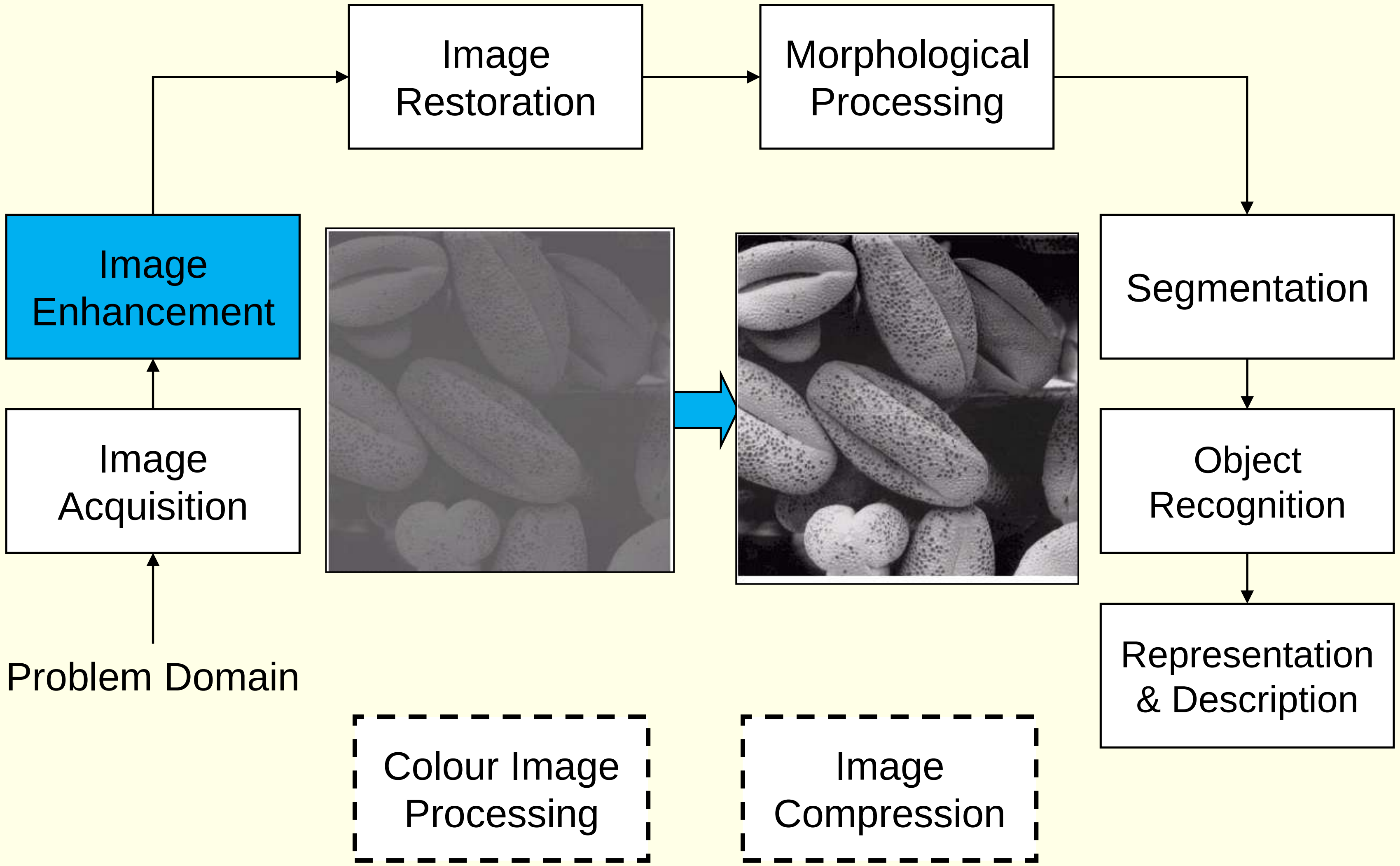
# Key Stages in Digital Image Processing



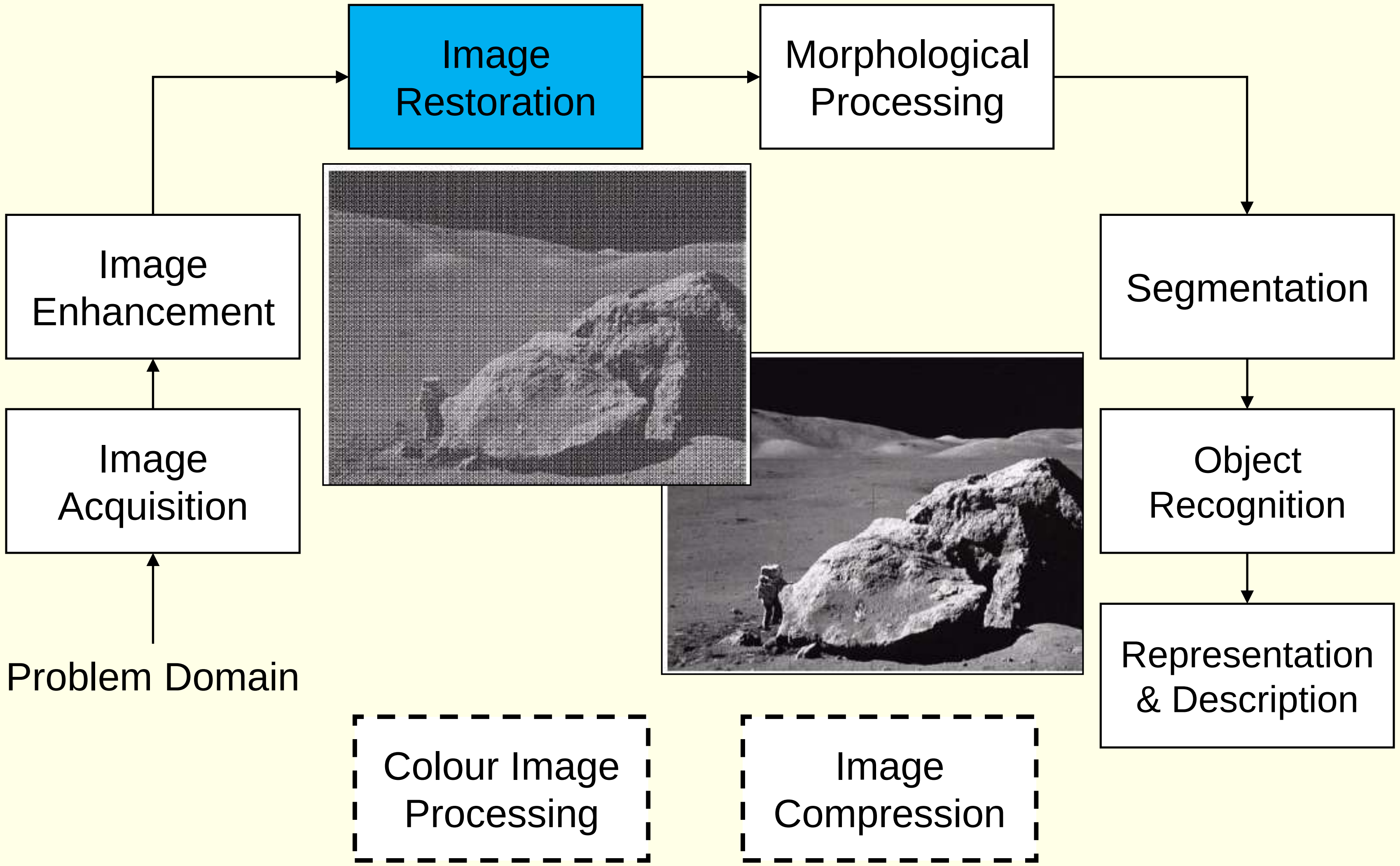
# Key Stages in Digital Image Processing:
# Image Acquisition



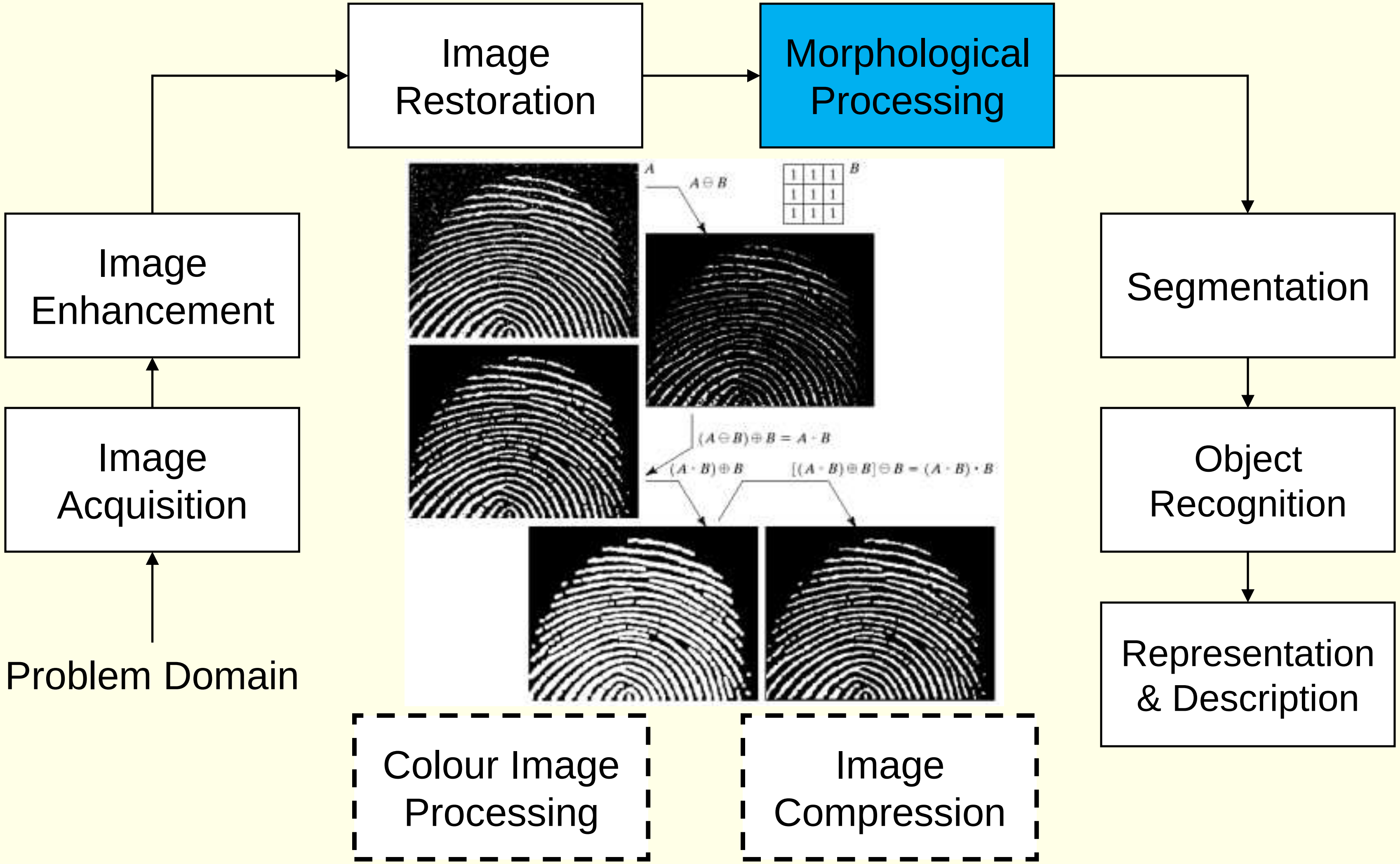
# Key Stages in Digital Image Processing: Image Enhancement



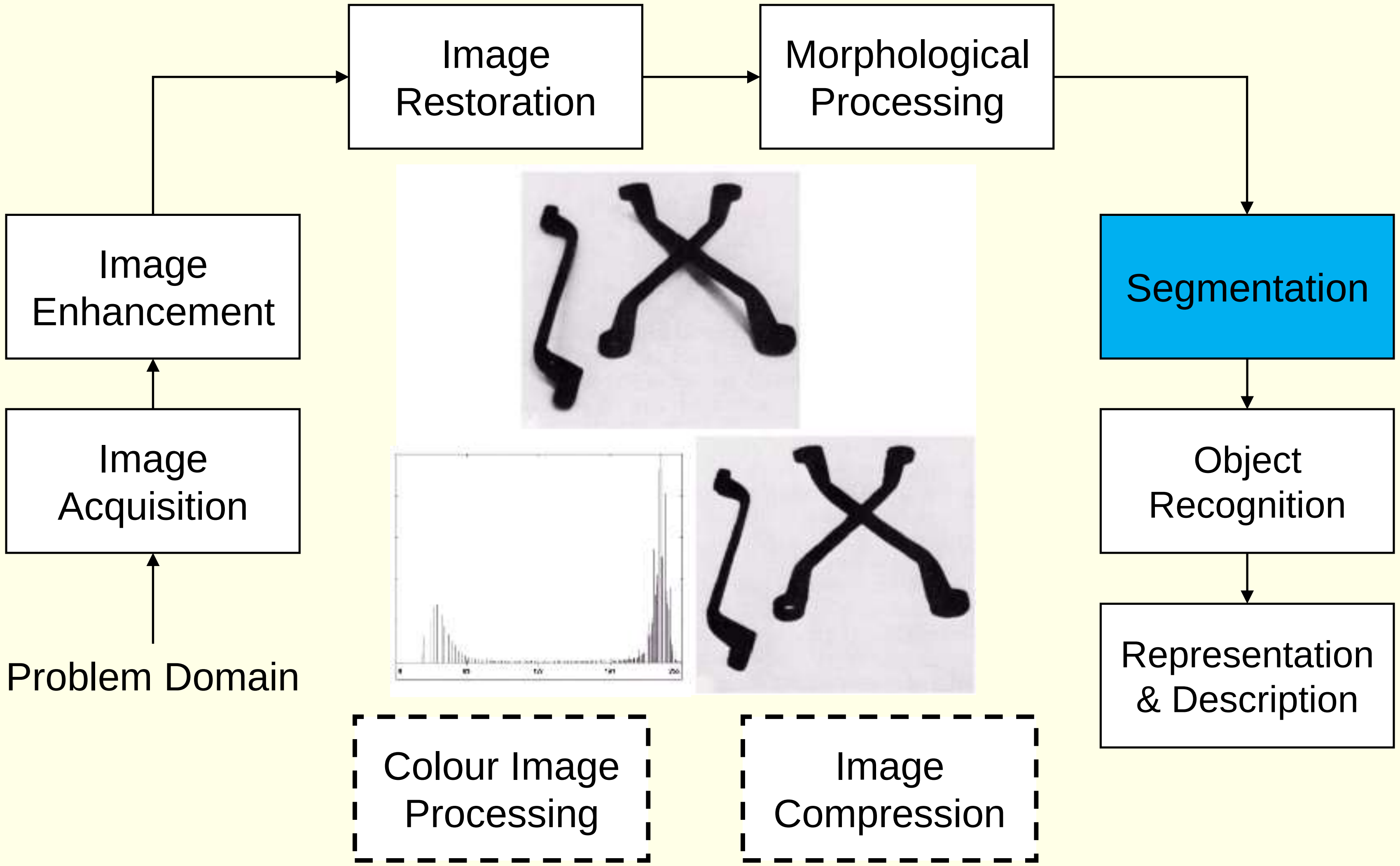
# Key Stages in Digital Image Processing: Image Restoration



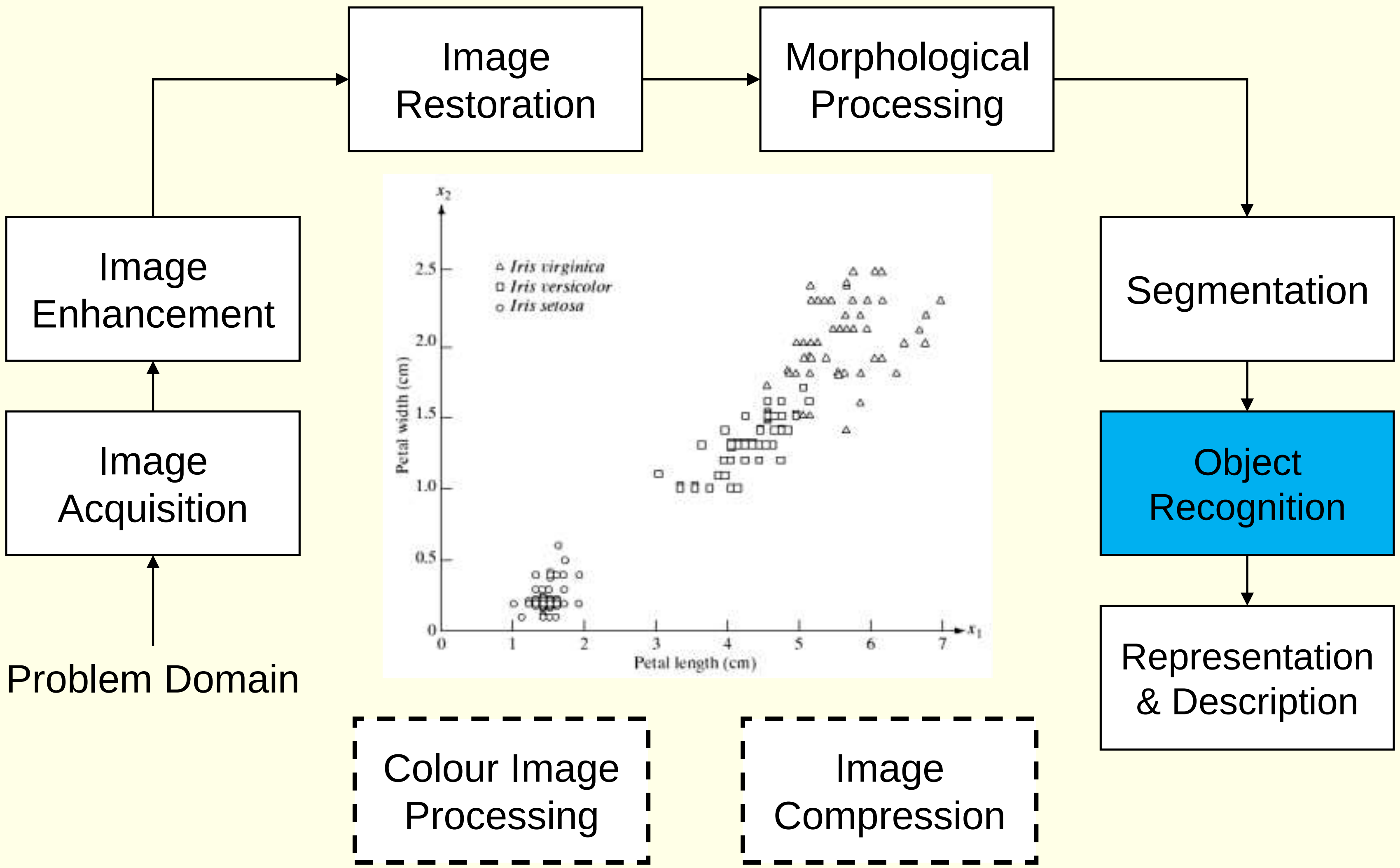
# Key Stages in Digital Image Processing: Morphological Processing



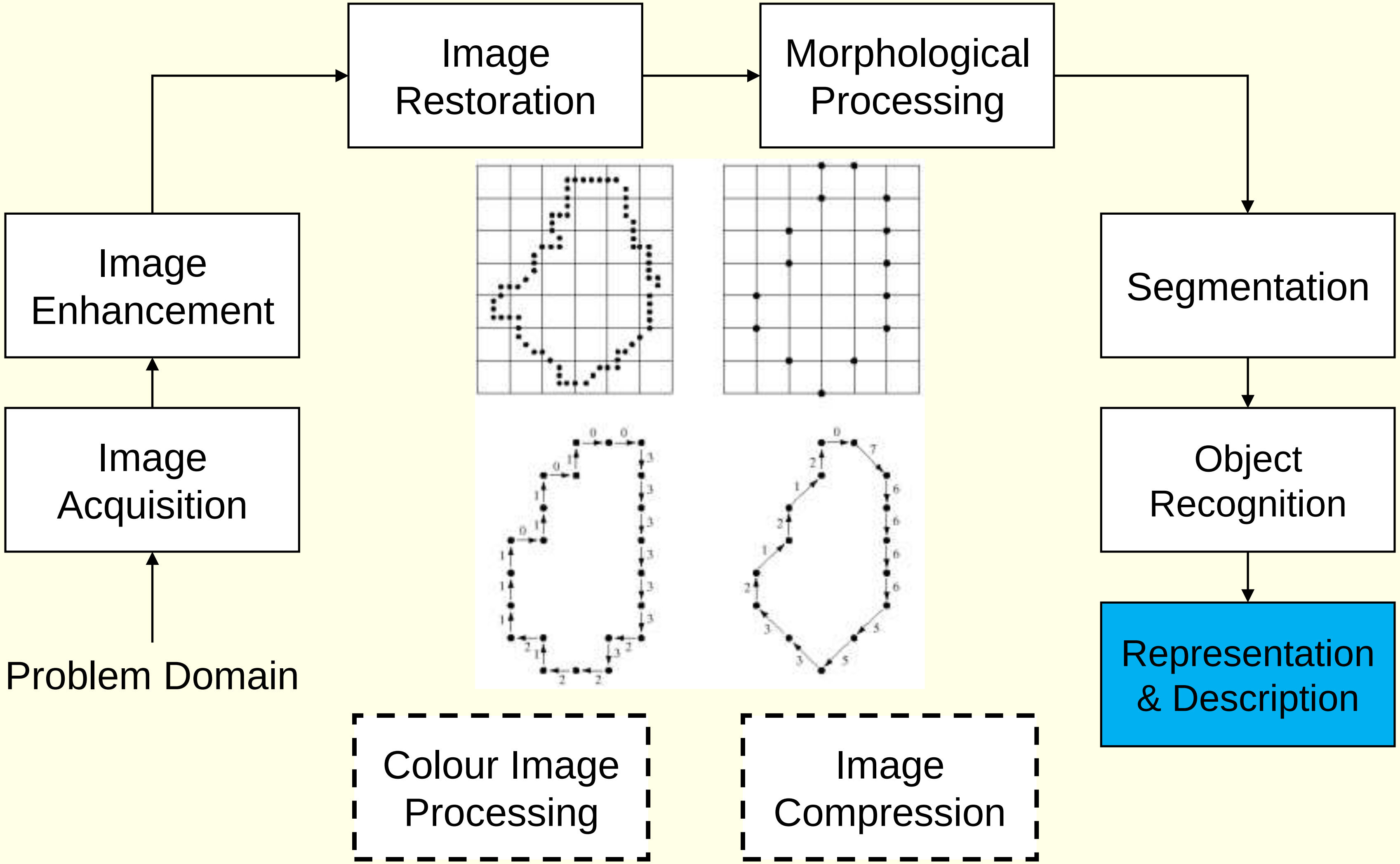
# Key Stages in Digital Image Processing: Segmentation



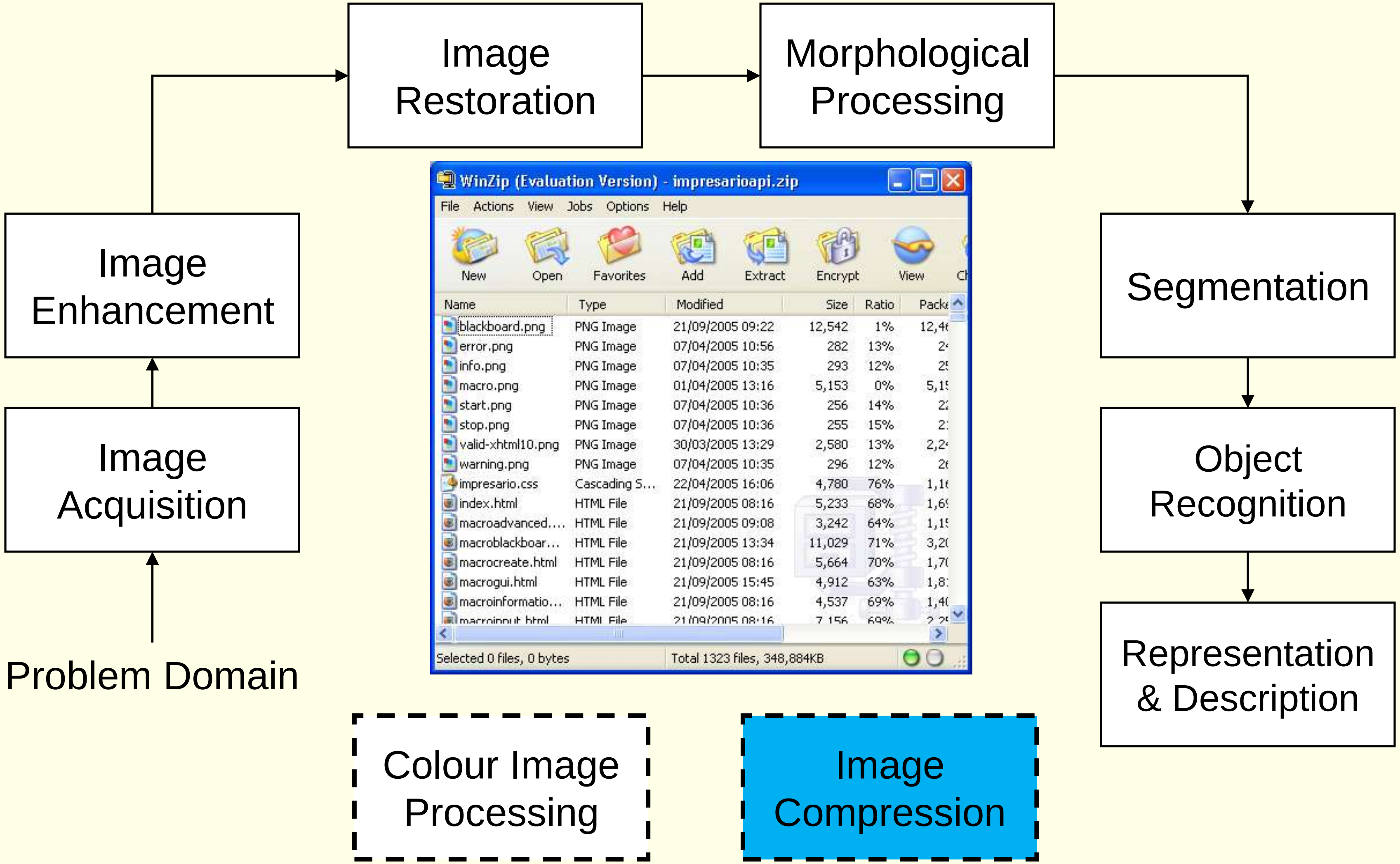
# Key Stages in Digital Image Processing: Object Recognition



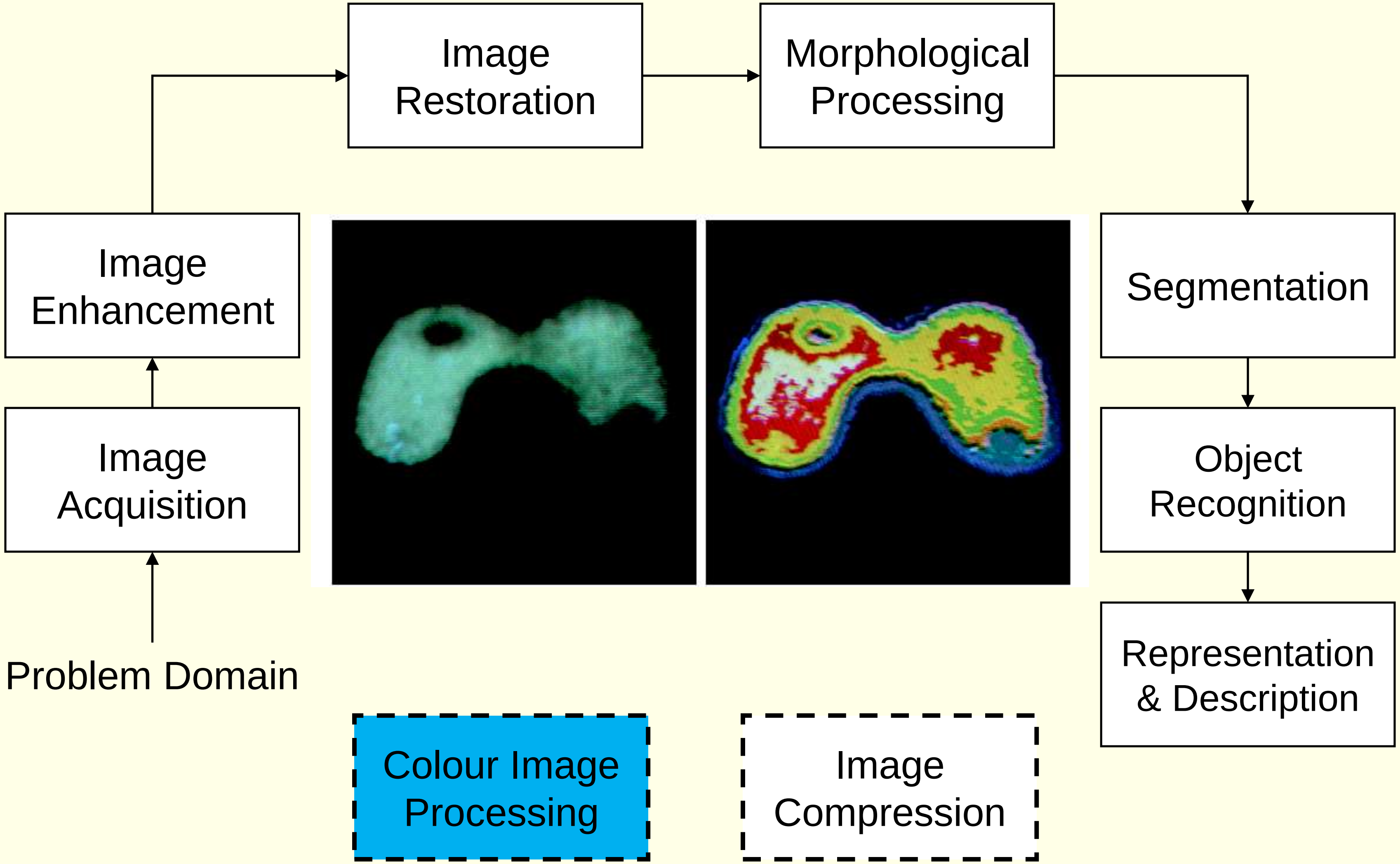
# Key Stages in Digital Image Processing: Representation & Description



# Key Stages in Digital Image Processing: Image Compression



# Key Stages in Digital Image Processing: Colour Image Processing



# Image Compression

# Elementary Information Theory

- How much information does a symbol convey?
- Intuitively, the more unpredictable or surprising it is, the more information is conveyed.
- Conversely, if we strongly expected something, and it occurs, we have not learnt very much.
- If p is the probability that a symbol will occur
- Then the amount of information, I, conveyed is:

$$\boxed{I = \log_2\left(\frac{1}{p}\right)}$$

- The information, I, is measured in bits
- It is the optimum code length for the symbol

# Elementary Information Theory

- The entropy, H, is the average information per symbol

$$H = \sum_s p(s)\log_2(\frac{1}{p(s)})$$

- Provides a lower bound on the compression that can be achieved.
- A simple example: Suppose we need to transmit four possible weather conditions:

   1. Sunny
   2. Cloudy
   3. Rainy
   4. Snowy

- If all conditions are equally likely, p(s)=0.25, and H=2
   - i.e. we need a minimum of 2 bits per symbol

# Lossless Compression coding

- Lossless coding means that
  - we are able to recover the information exactly

- Lossless compression techniques:
  - Variable length code words
  - Huffman code – integer code lengths
  - Arithmetic codes – non-integer code lengths

# Elementary Information Theory

- Huffman code

| Weather | Probability | Information | Integer code |
|---|---|---|---|
| Sunny | 0.5 | 1 | 0 |
| Cloudy | 0.25 | 2 | 10 |
| Rainy | 0.125 | 3 | 110 |
| Snowy | 0.125 | 3 | 111 |

- Previous illustration is an example of a lossless code
  - I.e. we are able to recover the information exactly

# Quantization

- Quantization is the process of approximating a continuous (or range of values) by a (much) smaller range of values.

$$\boxed{Q(x,\Delta) = \mathrm{Round}\left(\frac{x+0.5}{\Delta}\right)}$$

- Where Round(y) rounds y to the nearest integer
- $\Delta$ is the quantization step size

- Quantization plays an important role in lossy compression
  - This is where the loss happens

# Fundamentals of images

- An image consists of pixels (picture elements)
- Each pixel represents luminance (and colour)
  - Typically, 8-bits per pixel
- Colour
  - Colour spaces (representations)
    - RGB (red-green-blue)
    - CMY (cyan-magenta-yellow)
    - YUV
      - $Y = 0.3R+0.6G+0.1B$ (luminance)
      - $U=R-Y$
      - $V=B-Y$
- Greyscale
- Binary

# Fundamentals of images

- A TV frame is about 640x480 pixels
- If each pixels is represented by 8-bits for each colour, then the total image size is
  - $640 \times 480 * 3 = 921{,}600$ bytes or $\approx 7.4$Mbits
- At 30 frames per second, this would be $\approx$ 220Mbits/second
- Do we need all these bits????

Lossless compressed image, ~150KB

JPEG compressed, ~14KB

# Fundamentals of images

- Do we need all these bits?
  - No!
- The previous example illustrated the eye's sensitivity insensitive to high spatial frequencies

# 2-D DCT Transform

- Let i(x,y) represent an image with N rows and M columns
- Its DCT I(u,v) is given by

$$I(u,v) = \frac{1}{4} C(u)C(v)\left[\sum_{x=1}^{M}\sum_{y=1}^{N} i(x,y)\cos\left(\frac{(2x+1)u\pi}{16}\right)\cos\left(\frac{(2y+1)v\pi}{16}\right)\right]$$

- where

$$C(0) = \frac{1}{\sqrt{2}}$$

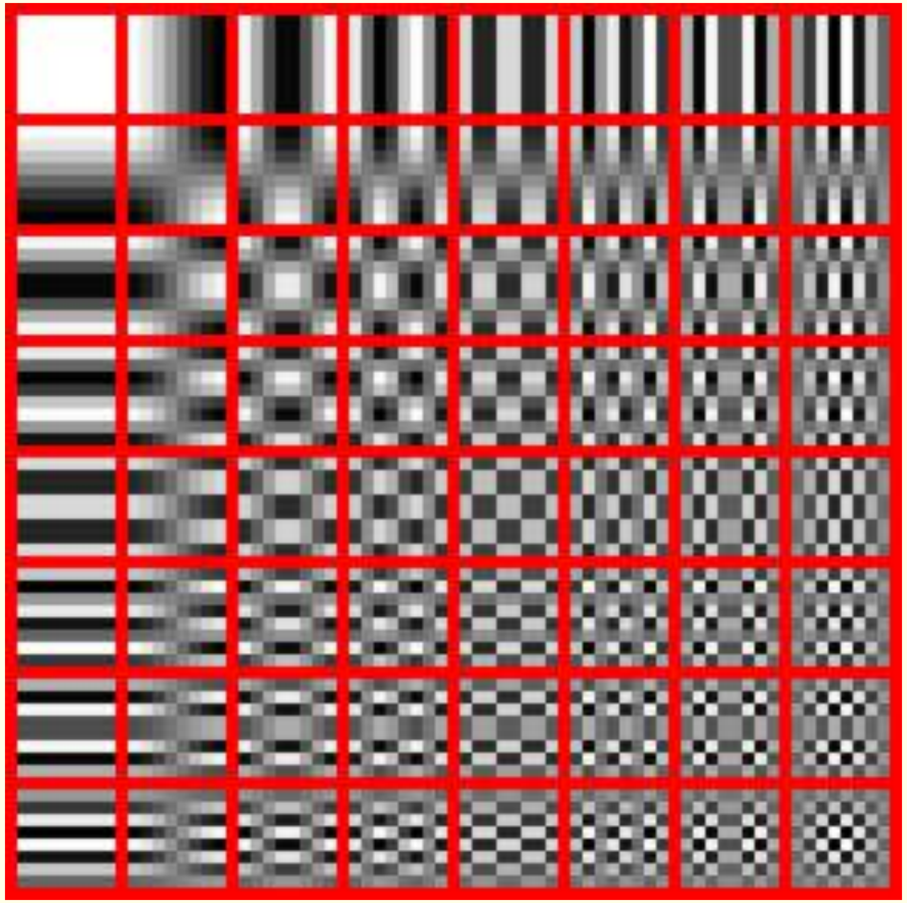
$$C(u) = 1$$

# Fundamentals of images

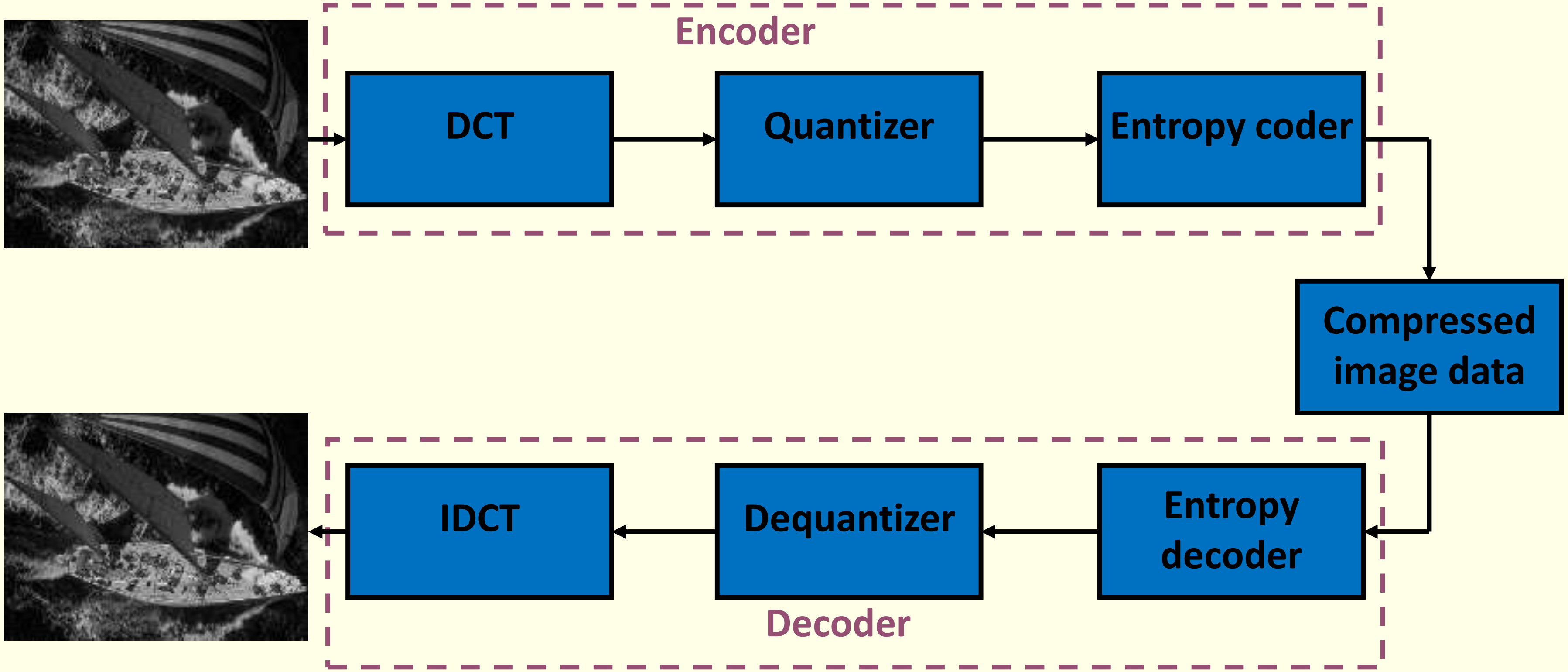
- Discrete cosine transform
  - Coefficients are approximately uncorrelated
    - Except DC term
    - C.f. original 8×8 pixel block
  - Concentrates more power in the low frequency coefficients
  - Computationally efficient
- Block-based DCT
  - Compute DCT on 8×8 blocks of pixels

# Fundamentals of images

- Basis functions for the 8×8 DCT (courtesy Wikipedia)



# Fundamentals of JPEG



# Fundamentals of JPEG

- JPEG works on 8×8 blocks
- Extract 8×8 block of pixels
- Convert to DCT domain
- Quantize each coefficient
  - Different step size for each coefficient
    - Based on sensitivity of human visual system
- Order coefficients in zig-zag order
- Entropy code the quantized values

# Fundamentals of JPEG

- A common quantization table is

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 16 | 11 | 10 | 16 | 24 | 40 | 51 | 61 |
| 12 | 12 | 14 | 19 | 26 | 58 | 60 | 55 |
| 14 | 13 | 16 | 24 | 40 | 57 | 69 | 56 |
| 14 | 17 | 22 | 29 | 51 | 87 | 80 | 62 |
| 18 | 22 | 37 | 56 | 68 | 109 | 103 | 77 |
| 24 | 35 | 55 | 64 | 81 | 104 | 113 | 92 |
| 49 | 64 | 78 | 87 | 103 | 121 | 120 | 101 |
| 72 | 92 | 95 | 98 | 112 | 100 | 103 | 99 |

# Fundamentals of JPEG

- Zig-zag ordering

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 5 | 6 | 14 | 15 | 27 | 28 |
| 2 | 4 | 7 | 13 | 16 | 26 | 29 | 42 |
| 3 | 8 | 12 | 17 | 25 | 30 | 41 | 43 |
| 9 | 11 | 18 | 24 | 31 | 40 | 44 | 53 |
| 10 | 19 | 23 | 32 | 39 | 45 | 52 | 54 |
| 20 | 22 | 33 | 38 | 46 | 51 | 55 | 60 |
| 21 | 34 | 37 | 47 | 50 | 56 | 59 | 61 |
| 35 | 36 | 48 | 49 | 57 | 58 | 62 | 63 |

# Fundamentals of JPEG

- Entropy coding
    - Run length encoding followed by
        - Huffman
        - Arithmetic
- DC term treated separately
    - Differential Pulse Code Modulation (DPCM)
- 2-step process
    1. Convert zig-zag sequence to a symbol sequence
    2. Convert symbols to a data stream

# Examples of varying JPEG compression ratios

500KB image, min. compression

40KB image, half compression

11KB image, max compression

# Wavelet vs. JPEG compression

**Wavelet compression**
**file size: 1861 bytes**
**compression ratio - 105.6**

**JPEG compression**
**file size: 1895 bytes**
**compression ratio - 103.8**

Source: “About Wavelet Compression”. http://www.barrt.ru/parshukov/about.htm.


# Thank you for your attention